



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 98347

To: Daniel Sullivan
Location: CM1-11E12
Art Unit: 1636
Thursday, July 17, 2003

Case Serial Number: 09/914191

From: Beverly Shears
Location: Biotech-Chem Library
CM1-1E05
Phone: 308-4994

beverly.shears@uspto.gov

Search Notes

Daniel,

Due to seq. length, unable to search amino acid query. I did, however, search Seq. ID 1 in the AA dBs.

Beverly

Note: Qy has an N in seq so 99.8% = 100% match

THIS PAGE BLANK (USPTO)

From: Sullivan, Daniel
Sent: Tuesday, July 08, 2003 4:18 PM
To: STIC-Biotech/ChemLib
Subject: Search Request 09914191

Please search for the following in the commercial and issued patent databases:

-a nucleic acid comprising SEQ ID NO:1

Also, I have reason to believe that the reverse complement of SEQ ID NO:1 encodes all or a portion of the following amino acid sequence:

MTHVASQFASSYVIFYWRDYFEDQPLL YPPGFDGRVVVYPSNQTLKDYLSWRQADCHINNLYNTVFW
ALIQQSGLTPVQAQGR LQGT LAADKNEILFSEFNINYNNELPMYRKGT VLIWQKVDEVMTKEIKLPTE
MEGKKMAVTRTRTKPVPLHCDIIGDAFWKEHPEILDEDS

would it be possible to search this amino acid sequence against the commercial and issued patent databases as well?

Thanks,
Dan

Daniel M. Sullivan
Examiner AU 1636
Room: 12D12
Mail Box: 11E12
Tel: 703-305-4448

RECEIVED
JUL - 8 2003
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Requestor's
Name: _____

Serial
Number: _____

Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 07-16-03

Searcher: Beverly 24994

Terminal time: 20

Elapsed time: _____

CPU time: _____

Total time: 25

Number of Searches: _____

Number of Databases: 1

Search Site

☐ STIC

☐ CM-1

☐ Pre-S

Type of Search

☐ N.A. Sequence

☐ A.A. Sequence

☐ Structure

☐ Bibliographic

Vendors

☐ IG

☐ STN

☐ Dialog

☐ APS

☐ Geninfo

☐ SDC

☐ DARC/Questel

☒ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 9, 2003, 12:44:53 ; Search time 38.5 Seconds
(without alignments)
4139.422 Million cell updates/sec

Title: US-09-914-191-1
Perfect score: 1096
Sequence: 1 ttggaatagttcttcttta.....gggttagtcacagattgttg 598

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgr2_1/USPTO.spool/US09914191/runat_09072003_112348_12858/app.query.fasta_1.775
-DB=A_Geneseq_101002 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOCLIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09914191 @CGN 1.1 114 @runat_09072003_112348_12858 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGIOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	544	49.2	173	22	AAB94556	Human protein sequ
C 2	544	49.2	298	22	AAM38811	Human polypeptide
C 3	537	48.6	298	22	AAB93770	Human protein sequ
C 4	455	41.2	85	22	AAG74471	Human colon cancer
C 5	427	38.6	297	22	AAM40597	Human polypeptide
C 6	219	19.8	286	22	ABB60415	Drosophila melanog
C 7	129.5	11.7	497	21	ABG39161	Arabidopsis thalia
C 8	129.5	11.7	502	21	ABG39160	Arabidopsis thalia
C 9	129.5	11.7	520	21	ABG39159	Arabidopsis thalia
C 10	128	11.6	470	21	AAG48957	Arabidopsis thalia
C 11	128	11.6	505	21	AAG48956	Arabidopsis thalia
C 12	128	11.6	523	21	AAG48955	Arabidopsis thalia
C 13	90.5	8.3	694	23	ABE92421	Herbicidally activ
C 14	88.5	8.1	128	21	AAG58628	Arabidopsis thalia
C 15	81.5	7.4	982	19	AAW77288	Zebraphish differ
C 16	80	7.3	233	21	AAB56951	Human prostate can
C 17	80	7.3	673	21	AAV96449	Forhead transcript
C 18	79	7.2	224	22	ABG03256	Novel human diagno
C 19	78	7.1	743	22	ABG09442	Novel human diagno
C 20	78	7.1	833	23	AAM47757	Mature EAI protein
C 21	77.5	7.0	277	20	AAV37752	Amino acid sequenc
C 22	76	6.9	419	22	AAG98793	Human cell death p
C 23	76	6.9	482	22	AAU27892	Human cell death p
C 24	76	6.9	488	22	AAG98787	Human cell death p
C 25	76	6.9	490	21	AAV49289	Mouse Glc1A polype
C 26	76	6.9	490	22	AAG98786	Human cell death p
C 27	76	6.9	578	22	AAU27720	Human full-length
C 28	76	6.9	580	21	AAV93274	Amino acid sequenc
C 29	76	6.9	625	22	ABE95823	Human protein sequ
C 30	76	6.9	1167	22	ABE04310	Human cardiac aden
C 31	75.5	6.8	393	17	AAW04404	Human CRTAM. Homo
C 32	75.5	6.9	604	23	ABP41697	Human ovarian anti
C 33	75.5	6.9	987	22	AAM39227	Human polypeptide
C 34	75.5	6.9	1086	22	AAM41013	Human polypeptide
C 35	75	6.8	227	22	ABG14493	Novel human diagno
C 36	75	6.8	393	21	AAE57000	Human prostate can
C 37	75	6.8	1195	22	AAU43884	Propionibacterium
C 38	75	6.8	1583	22	ABG09233	Novel human diagno
C 39	75	6.8	1583	22	ABG30211	Novel human diagno
C 40	74.5	6.7	333	23	ABG61771	Novel cathepsin-L
C 41	74.5	6.8	416	22	AAU32725	Novel human secret
C 42	74.5	6.8	505	22	ABE71148	Drosophila melanog
C 43	74.5	6.7	750	21	AAV92831	Mutant human prost
C 44	74	6.7	202	12	AAV10744	Non-A non-B hepati
C 45	74	6.8	240	21	AAG27963	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB94556
ID AAB94556 standard; Protein; 173 AA.
XX
AC AAB94556;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:15324.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX PR Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 15324; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 173 AA;

Alignment Scores:
 Pred. No.: 4, 86e-52 Length: 173
 Score: 544.00 Matches: 103
 Percent Similarity: 99.04% Conservative: 0
 Best Local Similarity: 99.04% Mismatches: 1
 Query Match: 49.23% Indels: 0
 DB: 22 Gaps: 0

US-09-914-191-1 (1-598) x AAB94556 (1-173)

QY 598 CACAACTCTGGACTTAACACACAGTACAGCCCAAGGAGGATTACAGGAACTCTTCAGCA 539
 DB 70 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaA 89
 QY 538 GACAAGAAATCAGATTTGTTTCTGTAATTCACATCAACTATATATGAGTGCAGATG 479
 DB 90 AspLysAsnGluLeuLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluLeuProMet 109
 QY 478 TATAGGAAGGGACTGTGTTGATATATGCGAAGAGTGGATGATGACAAAGAAAT 419
 DB 110 TyrArgLysGlyThrValLeuIleTgPInLysValAspGluValMetThrLysGluIle 129
 QY 418 AAGCTGCCACAGAAATGGAGAAAACATGCGACTGACCCGACCGACGACCAAGGCCA 359
 DB 130 LysLeuProThrGluMetGluGlyLysLysMetAlaValThrArgThrArgThrLysPro 149
 QY 358 GTGCCCTTGCACATCGCATATATCATCGGGGATGCTTTCTTGGAAAGGAACATCCAGAGATTCTA 299

Db 150 ValProLeuHisCysAspIlelleGlyAspAlaPheTrpLysGluHisProGluIleLeu 169
 QY 298 GATGAAGACAGC 287
 Db 170 AspGluAspSer 173
 RESULT 2
 AAM38811
 ID AAM38811 standard; Protein; 298 AA.
 XX AAM38811;
 XX 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 1956.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 XX N-PSDB; AAI57967.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 3; SEQ ID NO 1956; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX SQ Sequence 298 AA;

Alignment Scores:

Pred. No.: 5,7e-52 Length: 298
 Score: 544.00 Matches: 103
 Percent Similarity: 99.04% Conservative: 0
 Best Local Similarity: 99.04% Mismatches: 1
 Query Match: 49.23% Indels: 0
 DB: 22 Gaps: 0

US-09-914-191-1 (1-598) x AAB38811 (1-298)

QY 598 CACAACTCTGACTACACACAGTACAGCCCAAGGGAGATTACAGGAACTCTTNCAGCA 539
 |||||
 Db 195 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaA 214
 |||||
 QY 538 GACAAGAATCAGATTGTTCTTGAATTCACATCAACTATAATATGAGTGCAGTG 479
 |||||
 Db 215 AspLysAsnGluLeuLeuPheSerGluPheAsnLeuAsnTyrAsnAsnGluLeuProMet 234
 |||||
 QY 478 TATAGGAAGGACTGTGTTGATATGTCAGAAGGTGGATGAGTATGACAAAGAAATT 419
 |||||
 Db 235 TyrArgLysGlyThrValLeuLeuTyrGlnLysValAspGluValMetThrLysGluLe 254
 |||||
 QY 418 AACTGCGCAACAGAAATGGAAGGAAAAAGATGCGAGTACCCGGACCGACCAAGCCCA 359
 |||||
 Db 255 LysLeuProThrGluMetGluGlyLysLysMetAlaValThrArgThrLysPro 274
 |||||
 QY 358 GTCCCTTGCACGTGCATATCATCGGGGATGCTTCTTGAAGGAACATCCAGAGATTCTA 299
 |||||
 Db 275 ValProLeuHisCysAspIleIleGlyAspAlaPheTrpLysGluHisProGluLeuLe 294
 |||||
 QY 298 GATGAAGACAGC 287
 |||||
 Db 295 AspGluAspSer 298

RESULT 3

AAB93770

ID AAB93770 standard; Protein; 298 AA.

AC AAB93770;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:13475.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

OS EP1074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

PS Claim 8; SEQ ID 13475; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 298 AA;

Alignment Scores:

Pred. No.: 3,46e-51 Length: 298
 Score: 537.00 Matches: 102
 Percent Similarity: 98.08% Conservative: 0
 Best Local Similarity: 98.08% Mismatches: 2
 Query Match: 48.60% Indels: 0
 DB: 22 Gaps: 0

US-09-914-191-1 (1-598) x AAB93770 (1-298)

QY 598 CACAACTCTGACTACACACAGTACAGCCCAAGGGAGATTACAGGAACTCTTNCAGCA 539
 |||||
 Db 195 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaA 214
 |||||
 QY 538 GACAAGAATCAGATTGTTCTTGAATTCACATCAACTATAATATGAGTGCAGTG 479
 |||||
 Db 215 AspLysAsnGluLeuLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluProMet 234
 |||||
 QY 478 TATAGGAAGGACTGTGTTGATATGTCAGAAGGTGGATGAGTATGACAAAGAAATT 419
 |||||
 Db 235 TyrArgLysGlyThrValLeuLeuTyrGlnLysValAspGluValMetThrLysGluLe 254
 |||||
 QY 418 AACTGCGCAACAGAAATGGAAGGAAAAAGATGCGAGTACCCGGACCGACCAAGCCCA 359
 |||||
 Db 255 LysLeuProThrGluMetGluGlyLysLysMetAlaValThrArgThrLysPro 274
 |||||
 QY 358 GTCCCTTGCACGTGCATATCATCGGGGATGCTTCTTGAAGGAACATCCAGAGATTCTA 299
 |||||
 Db 275 ValProLeuHisCysAspIleIleGlyAspAlaPheTrpLysGluHisProGluLeuLe 294
 |||||
 QY 298 GATGAAGACAGC 287
 |||||
 Db 295 AspGluAspSer 298

RESULT 4

AAG74471

ID AAG74471 standard; Protein; 85 AA.

XX AAG74471;

AC AAG74471;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:5235.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma.

XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX DR N-PSDB; AAH33902.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 6911-6912; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated P,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 85 AA;

Alignment Scores:
Pred. No.: 3,55e-42 Length: 85
Score: 455.00 Matches: 85
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.18% Indels: 0
DB: 22 Gaps: 0

US-09-914-191-1 (1-598) x AAG74471 (1-85)
QY 541 GCACAGCATGAGATTTGTTCTGATTCACATCACTAATATGATGCTGCCG 482
DB 1 ALaAspLysAsnGluLeuLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluLeuPro 20
QY 481 ATGTATAGCAAGGCGACTGTGTTGATATGCGAAGGTGATGATGATGACAAAGAA 422
DB 21 MetTyrArgLysGlyThrValLeuIleTrpGlnLysValAspGluValMetThrLysGlu 40
QY 421 ATTAAGCTGCCACAGAAATGGAAGGAAAAAGATGCGAGTGCACCGGACCGAGCAAG 362
DB 41 IleLysLeuProThrGluMetGluGlyLysLysMetAlaValThrArgThrLys 60
QY 361 CCAGTCCCTTGCACATCATCGGGATCTTCTTGGAGGACATCATCAGAGATT 302
DB 61 ProValProLeuHisCysAspIleIleGlyAspAlaPheTrpLysGluHisProGluIle 80
QY 301 CTAGATGAACACAGC 287

Db 81 LeuAspGluAspSer 85
RESULT 5
AAM40597
ID AAM40597 standard; Protein; 297 AA.
XX AC AAM40597;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 5528.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX DR N-PSDB; AAI59753.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX Example 2; SEQ ID NO 5528; 10078pp; English.
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
SQ Sequence 297 AA;

Alignment Scores:
Pred. No.: 6,94e-39 Length: 297
Score: 427.00 Matches: 89

	Percent Similarity:	67.42%	Conservative:	0
	Best Local Similarity:	67.42%	Mismatches:	5
	Query Match:	38.64%	Indels:	38
	DB:	22	Gaps:	2
US-09-914-191-1 (1-598) x AAM40597 (1-297)				
QY	598	CACAACTCGGACTTAACACAGCAGTACCAAGCCCAAGGGAGATTACAGGGAATCTTTNCAGCA	539	
Db	203	GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaAla	222	
QY	538	GACAAGAATCAGATTTGTTCCTGAATTCACCATCAACTATAATATGAGCTGCCGATG	479	
Db	223	AspLysAsnGluIleLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluProMet	242	
QY	478	TATAGAAAAGGAGCTGTGTTGCATATGCAGAGGTGGATGAAGTGATGACAAAGAAAT	419	
Db	243	TyrArgLysGlyThrValLeuIleTrpGlnLysValaspGluValMetThrLysGluIle	262	
QY	418	AAGCTGCCAACAGAAATGGAAGAAAAAGATGGCAGTGACCCGGACCAGGACAAAGCCA	359	
Db	263	LysLeuProThrGluMetGluGlyLysLysMetAlaValThrArgThrArgThrLysPro	282	
QY	358	GTG---CCCTTGCACTCGGATATCATCGGGGATGCTTCTGGNAGGAACATCCAGAGATT	302	
Db	283	CysLysProSerHisLeu-----	288	
QY	301	CTAGATGAACACAGCTGACCCCTTTTGGCTTCAGTTCTGGTGCTTAACCATGCAAGCC	242	
Db	288	-----	288	
QY	241	CTCCCACCTCCCGGGCTCCTTCCTTAGTGCGTG	206	
Db	289	-----ProArgAlaProCysLeuArgTipLeu	297	
RESULT 6				
ABB60415	ID	ABB60415 standard; Protein; 286 AA.		
XX	AC	ABB60415;		
XX	DT	26-MAR-2002 (first entry)		
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 8037.		
KW	KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.		
OS	OS	Drosophila melanogaster.		
PN	PN	WO200171042-A2.		
XX	PD	27-SEP-2001.		
XX	PF	23-MAR-2001; 2001WO-US09231.		
PR	PR	23-MAR-2000; 2000US-191637P.		
PR	PR	11-JUL-2000; 2000US-0614150.		
XX	PA	(PEXE) PE CORP NY.		
XX	Venter JC,	Adams M, Li PWD, Myers EW;		
PI	PI	WI; 2001-656860/75.		
DR	DR	N-FSDB; ABL04518.		
XX	XX	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -		
XX	PS	Disclosure; SEQ ID NO 8037; 2lpp + Sequence Listing; English.		
XX	CC	The invention relates to an isolated nucleic acid detection reagent		

capable of detecting 1000 or more genes from Drosophila. The invention useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABLJ30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).				
The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.				
SQ	Sequence	286 AA;		
Alignment Scores:				
Pred. No.:	1.26e-15	Length:	286	
Score:	219.00	Matches:	47	
Percent Similarity:	60.61%	Conservative:	13	
Best Local Similarity:	47.47%	Mismatches:	21	
Query Match:	19.82%	Indels:	18	
DB:	22	Gaps:	2	
US-09-914-191-1 (1-598) x ABB60415 (1-286)				
QY	595	CAATCTGGACTTAACACAGTACCAAGCCCAAGGGAGATTACAGGGAATCTTTNCAGCAGAC	536	
Db	166	GluLysGlyLeuThrAsnGlnGlnAlaGluAlaLysLeuArgGlyThrPheSerAlaasp	185	
QY	535	AAGATGAGATTTTGTTCCTGAATTCACCATCAACTATATATGAGTCCGCGATGAT	476	
Db	186	LysAsnGluLeuPheGlnGluPheGlyIleAsnTyrAsnAsnLeuProAlaMetTyr	208	
QY	475	AGGAAAGGGAGCTGTGTTGCATATGCAGAGGTGGATGAAGTGATGACAAAAGAAATTAAG	416	
Db	206	ArgLysGlyThrIleLeuLeuArgLysArgVal-----	216	
QY	415	CTGCCAACAGAAATGGAAGGAAAGAAAGATGGCAGTGACCCGGACCAGGACAAAGCCAGTG	358	
Db	217	-----ileLeuGlyGluLys-----SerArgGlnAlaValVal	222	
QY	355	CCCTTGCACTCGGATATCATCGGGGATGCTTCTGGAAGGAAACATCCAGAGATTCTA	299	
Db	228	ProLeuHisGluAspLeuIleSerSerGlnPheTrpLysGluHisThrGluIleLeu	246	
RESULT 7				
AAG39161	ID	AAG39161 standard; Protein; 497 AA.		
XX	AC	AAG39161;		
XX	DT	18-OCT-2000 (first entry)		
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 48413.		
KW	KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.		
OS	OS	Arabidopsis thaliana.		
XX	PN	EP1033405-A2.		
XX	PD	06-SEP-2000.		
XX	PF	25-FEB-2000; 2000EP-0301439.		
XX	PR	25-FEB-1999; 99US-0121825.		
PR	PR	05-MAR-1999; 99US-0123180.		
PR	PR	09-MAR-1999; 99US-0123548.		
PR	PR	23-MAR-1999; 99US-0125788.		
PR	PR	25-MAR-1999; 99US-0126264.		
PR	PR	29-MAR-1999; 99US-0126785.		
PR	PR	01-APR-1999; 99US-0127462.		
PR	PR	06-APR-1999; 99US-0128234.		

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130891.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

Db 220 PheValHisSerGluAsnIleAlaGlyArgSerPheTrpAsnGluGlnProSerLeuTyr 239
QY 298 GATGAA 293
Db : : : : :
240 AsnAep 241
RESULT 9
AAG39159
ID AAG39159 standard; Protein; 520 AA.
XX
AC AAG39159;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48411.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-01611359.
 PR 26-OCT-1999; 99US-01611360.
 PR 26-OCT-1999; 99US-01611361.
 PR 28-OCT-1999; 99US-01611920.
 PR 28-OCT-1999; 99US-01611992.
 PR 28-OCT-1999; 99US-01611993.
 PR 28-OCT-1999; 99US-01611993.
 PR 29-OCT-1999; 99US-01612142.

Alignment Scores:

Pred. No.: 2,2e-05 Length: 470
 Score: 128.00 Matches: 32
 Percent Similarity: 56.57% Conservatives: 24
 Best Local Similarity: 32.32% Mismatches: 33
 Query Match: 11.58% Indels: 10
 DB: 21 Gaps: 3

US-09-914-191-1 (1-598) x AAG48957 (1-470)

QY 595 CAATCTGGACTACACCAAGCCAGGAGATTACAGGAACTCTTNCAGCAGAC 536
 Db 379 LysSerGlySerLysIleGlnAlaGlnAspTyrLeuLysGlyThrGlnThrArgGlu 398
 QY 535 AAGAATGAGATTGTGTTCTGAATCAACATCAACTATAATAATGAGCTGCCGATGAT 476
 Db 399 LysAsnGluLeuLeuSerGlnGlnPheGlyIleGluTyrAsnSerLeuProValIlePhe 418
 QY 475 AGAAAGGGACTGTGTGATATGCGAAGGTGGATGATGACAAAGAAATTAAG 416
 Db 419 ArgMetGlySerSerValPhe-----ArgLeuLysGluGlyValThrGluGlu----- 434
 QY 415 CTCCCAACAGAAATGGAAGAAAGATGCGCAGTGCACCGGACAGGACAAAGCCAGTG 356
 Db 435 ---AsnGlyGluValSerGlyLysGlnVal-----GluAlaGluValGly 448
 QY 355 CCCTTGCACTGCGATATCATCGGGGATGCTTTCTGGAAGGAACATCCAGAGATTCTA 299
 Db 449 ValAspTyrSerAsnIleIleAspGlnCysPheTrpGlnGlnHisProHisIleLeu 467

RESULT 11

AAG48956
 ID AAG48956 standard; Protein: 505 AA.

XX AC AAG48956;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 61881.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123348.

PR 25-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 08-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128314.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.	Alignment Scores:	
PR 05-AUG-1999;	99US-0147260.	Pred. No.:	2.25e-05
PR 06-AUG-1999;	99US-0147303.	Score:	128.00
PR 06-AUG-1999;	99US-0147416.	Percent Similarity:	56.57%
PR 09-AUG-1999;	99US-0147493.	Best Local Similarity:	32.32%
PR 09-AUG-1999;	99US-0147935.	Query Match:	11.58%
PR 10-AUG-1999;	99US-0148171.	DB:	21
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.	US-09-914-191-1 (1-598) x AAG48956 (1-505)	
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.	QY 595 CAATCTGGACTAACAACGACCAAGCCAGGAGATTACAGGAACTCTTNCAGCAGAC	536
PR 16-AUG-1999;	99US-0149175.	Db 414 LysSerGlyLysSerLysLysileGlnAlaGlnAspTyrLeuLysGlyThrGlnThrArgGlu	433
PR 17-AUG-1999;	99US-0149175.	QY 535 AAGAATGAGATTGTTGTTTCTGAATCAACATCAACTAATAATGAGTGCAGTGCAGTAT	476
PR 18-AUG-1999;	99US-0149426.	Db 434 LysAsnGluLeuLeuSerGlnGlnPheGlyileGluTyrAsnSerLeuProValIlePhe	453
PR 20-AUG-1999;	99US-0149722.	QY 475 AGGAAGGAGCTGTGTTGATATGGCAGAGGTGGATGAAGTATGATGACAAAGAAATTAAG	416
PR 20-AUG-1999;	99US-0149723.	Db 454 ArgMetGlySerSerValPhe	469
PR 20-AUG-1999;	99US-0149929.	QY 415 CTGCCAACAGAAATGAAGGAAAAAGATGGCAGTGACCCGACGAGCAAGCCAGTG	356
PR 23-AUG-1999;	99US-0149902.	Db 470 ---AsnGlyGluValSerGlyLysGlnVal	483
PR 23-AUG-1999;	99US-0149930.	QY 355 CCCTTGCACTGCGATATCATCGGGATGCTTTCTGGAAGGAACATCCAGAGATTCTA	299
PR 25-AUG-1999;	99US-0150566.	Db 484 ValAspTyrSerAsnIleileAspGlnCysPheTyrGlnGlnHisProHisIleLeu	502
PR 26-AUG-1999;	99US-0150884.	RESULT 12	
PR 27-AUG-1999;	99US-0151066.	AAG48955	
PR 27-AUG-1999;	99US-0151080.	ID AAG48955 standard; Protein; 523 AA.	
PR 30-AUG-1999;	99US-0151303.	XX	
PR 31-AUG-1999;	99US-0151438.	AC AAG48955;	
PR 01-SEP-1999;	99US-0151930.	XX	
PR 07-SEP-1999;	99US-0152363.	DT 18-OCT-2000 (first entry)	
PR 10-SEP-1999;	99US-0153070.	XX	
PR 13-SEP-1999;	99US-0153758.	XX	
PR 16-SEP-1999;	99US-0154018.	XX	
PR 18-SEP-1999;	99US-0154039.	XX	
PR 20-SEP-1999;	99US-0154779.	XX	
PR 22-SEP-1999;	99US-0155139.	XX	
PR 23-SEP-1999;	99US-0155486.	XX	
PR 24-SEP-1999;	99US-0155659.	XX	
PR 28-SEP-1999;	99US-0156458.	XX	
PR 29-SEP-1999;	99US-0156596.	XX	
PR 04-OCT-1999;	99US-0157117.	XX	
PR 05-OCT-1999;	99US-0157753.	XX	
PR 06-OCT-1999;	99US-0157865.	XX	
PR 07-OCT-1999;	99US-0158029.	XX	
PR 08-OCT-1999;	99US-0158232.	XX	
PR 12-OCT-1999;	99US-0158369.	XX	
PR 13-OCT-1999;	99US-0159293.	XX	
PR 13-OCT-1999;	99US-0159294.	XX	
PR 13-OCT-1999;	99US-0159329.	XX	
PR 14-OCT-1999;	99US-0159329.	XX	
PR 14-OCT-1999;	99US-0159330.	XX	
PR 14-OCT-1999;	99US-0159331.	XX	
PR 14-OCT-1999;	99US-0159637.	XX	

Arabidopsis thaliana protein fragment SEQ ID NO: 61880.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 03-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.


```
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 2,27e-05
Score: 128.00
Length: 523
Matches: 32
Percent Similarity: 56.57%
Conservative: 24
Best Local Similarity: 32.32%
Mismatch: 33
Query Match: 11.58%
Indels: 10
DB: 21
Gaps: 3

US-09-914-191-1 (1-598) x AAG48955 (1-523)
QY 595 CAATCTGGACTACACAGTACAGCCCAAGGAGGATTACAGGAACTCTTCAGCAGAC 536
DB 432 LysSerGlyLysSerLysIleGlnAlaGlnAspTyrLeuLysGlyThrGlnThrArgGlu 451
QY 535 AAGAATGAGATTGTTGTTCTGAATCAACATCAACTAATATATAGTGCCTGATGAT 476
DB 452 LysAsnGluLeuLeuSerGlnGlnPheGlyIleGluTyrAsnSerLeuProValIlePhe 471
QY 475 AGGAAAGGAGTGTGTTGATATGCGAGAGGTGGATGATGATGATGATGATGATGATGAT 416
DB 472 ArgMetGlySerValPhe-----ArgLeuLysGluGlyValThrGluGlu----- 487
QY 415 CTGCCACAGAAATGAGAAAGAAAGATGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 356
DB 488 ---AsnGlyGluValSerGlyLysGlnVal-----GluAlaGluValGly 501
QY 355 CCCTTGACGTCGATATCATCGGGATGCTTTCTGGAAGGAAACATCCAGAGATTCTA 299
DB 502 ValAspTyrSerAsnIleIleAspGlnCysPheTrpGlnGlnHisProHisIleLeu 520

RESULT 13
ID ABB92421
AC ABB92421;
XX
XX
XX 31-MAY-2002 (first entry)
```

```
XX Herbicidally active polypeptide SEQ ID NO 1632.
DE DE
XX Herbicidal; plant; agriculture; herbicide.
KW KW
XX Arabidopsis thaliana.
OS OS
XX WO200210210-A2.
PN PN
XX 07-FEB-2002.
PD PD
XX 28-AUG-2001; 2001WO-EP09892.
PF PF
XX 28-AUG-2001; 2001WO-EP09892.
PR PR
XX (FARB ) BAYER AG.
PA PA
XX Tietjen K, Weidler M;
PI PI
XX WPI; 2002-269010/31.
DR DR
XX
XX Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant from plant with nucleic acid or amino acid sequences from non-plant organisms -
PT PT
PT Claim 5; SEQ ID NO 1632; 261pp + Sequence Listing; English.
PS PS
XX The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.
CC CC
XX SQ Sequence 694 AA;
SQ SQ
Alignment Scores:
Pred. No.: 0.386 Length: 694
Score: 90.50 Matches: 33
Percent Similarity: 38.89% Conservative: 9
Best Local Similarity: 30.56% Mismatches: 46
Query Match: 8.26% Indels: 20
DB: 23 Gaps: 3

US-09-914-191-1 (1-598) x ABB92421 (1-694)
QY 74 CCAACACCCGCAACAGGTGTTCTGCTTAAACAGAGTAAAGATACACACCCCATCCATC 133
DB 160 ProProProProSerProAlaThrProSerThrProProArgSerProProPro----- 177
QY 134 CTTCTCTCCCTGTTCCCTCCCACTTGAGTTGTTGTCATTCGCACAGTGTCTGGTG 193
DB 178 ProSerThrProThrProProArg----- 186
QY 194 GTAGGATGCTCAGCAGCAGTAAAGCAAGAGCCCTGGAGGTGGAGGCTTGCATGTT 253
DB 187 ValGlySerLeuSerProProProProAlaSerPro-SerGlyGlyArgSerProSerTh 206
QY 254 TAAGCACACAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCACTTAGAATCTCTGGATG 313
DB 206 rProSerThrThrProGlySerSerProAlaGlnSerSerLysGluLeuSer----- 224
QY 314 TTCCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGAGGCAAGGCTTCTCTGTT 373
DB 225 -----LysGlyAlaMetValGlyIleAlaIleGlyGlyPheValLeuLe 240
QY 374 CCGGTCACATGCTCTTTT 395
DB 240 uValAlaLeuAlaLeuIlePhe 247
```

RESULT 14
AAG58628 standard; Protein; 128 AA.
ID AAG58628; AC AAG58628; DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75712.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121625.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159339.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.393
Score: 88.50
Percent Similarity: 41.90%
Best Local Similarity: 31.43%
Query Match: 8.07%
DB: 21

US-09-914-191-1 (1-598) x AAG58628 (1-128)
QY 85 AACAAAGTGTTCTCTTAAC-----AGAGTAAGATACACCCCCCATCAT 132
Db 3 SerlyslalleCysLeuGlyPheLeuProArgLeuArgPheSerProArgLeu 22
QY 133 CCCTTCCTCCCTGTTCTCCCTCCCACTTGAGTTGTGTCTTCGCCACCACTGCTGGGT 192

Db 23 ProSerLeuProSerProProAlaSerSerThrPheAlaThrArgHisLysLeuAsp 42
QY 193 GGTAGGATGCTACAGCCACCTAAGCAAGAGCCCTGGAGGTGGAGGCTTGATGG 252
Db 43 SerArgGlnThr-----LeuLeuTyr 49
QY 253 TTAAGCACACCAAGTAAGCGCAAAAGGTCAGTCTTTCATCTAGAACTCTGGAT 312
Db 50 ---AsnLysProGlnLeuSerArgGlyArgValalaCysSerSerGlnSerAspSer 68
QY 313 GTTCCTTCAGAAAGCATCCCGATGATATCGCAGTCAAGGCA---CTGGCTTTGTCC 369
Db 69 ArgProGluLysLysGlnSerAspLysSerAsnTyrAlaArgAlaGluLeuPheArgGly 88
QY 370 TGGTCGGGTCACTG 384
Db 89 LysSerGlySerVal 93

RESULT 15
AAW77288
ID AAW77288 standard; Protein; 982 AA.
XX
AC AAW77288;
XX
DT 20-NOV-1998 (first entry)
XX
DE Zebrafish differentiation enhancing factor 2 protein.
XX
KW Zebrafish; differentiation enhancing factor; ankyrin repeat; C2 domain;
KW SH3 consensus binding sequence; pleckstrin homology domain; adipogenesis;
XX neurogenesis; hyperplastic disease; neoplastic disease; nervous system.
XX Brachydanio rerio.
XX
PN WO9836065-A1.
XX
PD 20-AUG-1998.
XX
PF 13-FEB-1998; 98WO-US02724.
XX
PR 14-FEB-1997; 97US-0038191.
XX
PA (DAND ) DANA FARBER CANCER INST INC.
XX
PI Chan J, Harris DF, Hu E, King FJ, Spiegelman B;
PI Thomas RW;
XX
XX WPI; 1998-467173/40.
XX N-PSDB; AAV59105.
XX
PT New nucleic acid encoding differentiation enhancing factor - used
PT particularly to regulate adipogenesis and neurogenesis, e.g. for
PT treating tumours and neurological disease
XX
PS Claim 3; Fig 12; 203pp; English.
XX
CC The differentiation enhancing factors (DEF), comprise at least one each
CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
CC are mediators of SH3-domain dependent signalling and may be involved in
CC cellular gene expression, cytoskeletal architecture, protein trafficking,
CC endocytosis or adhesion, migration, proliferation and differentiation of
CC cells. Typical applications of DEF and agents that modulate
CC interaction between the protein and it's ligand, or of nucleic acid
CC expressing them, are treatment of hyperplastic and neoplastic disease
CC (a wide range of solid tumours and leukaemias), including metastases; for
CC in vitro induction of differentiation of neural crest cells to neurons,
CC glial cells etc.; for increasing neuron survival, and inducing cell
CC repair, in the nervous system (e.g. treatment of traumatic injury,
CC stroke, Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic
CC lateral sclerosis, multiple sclerosis etc.).
XX
```

SQ Sequence 982 AA;

Alignment Scores:

Pred. No.:	4.33	Length:	982
Score:	81.50	Matches:	33
Percent Similarity:	39.81%	Conservative:	10
Best Local Similarity:	30.56%	Mismatches:	38
Query Match:	7.38%	Indels:	27
DB:	19	Gaps:	5

us-09-914-191-1 (1-598) x AAW77288 (1-982)

QY	389	GATGGCAGTCACCGGACGAGCAAGCCAGTGCCTTCGACATGCGATATCATCGG---	333
DB	699	AspGluSerAspGluSerLeuAspGluSerProHisArgAspGluArgPro	718
QY	332	---GGATGCTTTCTGGAAGGAACA-----TCCAGA	306
DB	719	IleSerCysTyrThrProGlySerAsnSerLeuGlnLeuSerProAlaSerLeuSerArg	738
QY	305	GATTCTAGATGAACACAGCTGACCCCTTTTCGGCTTCAGTTCTCGT-----	261
DB	739	AspGlyArgAspLeuValLysAspLysGlnArgPheValProAsnLeuValAsnAsnGlu	758
QY	260	-----GTGCTTAACCATGCAAGCCCTCCACCTCCAGGCTCCTTGCCTTAG	213
DB	759	ThrTyrGlyThrIleIleAsnThrSerSerProValSerLeuSerSerSerAlaPro---	777
QY	212	GTGGCTGTAGCATCCCTTACCACCGACACATCGTGCGAATGACAACTCAAAGTTGGGA	153
DB	778	-----ProLeuProProArgAsnLeuValGlnProSerAlaLeu---AlaGly	792
QY	152	GGGGAACAGGGAAGGAGGATGG	129
DB	793	LeuThrGlnGlySerProGlyTrp	800

Search completed: July 9, 2003, 12:48:12
Job time : 44.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 9, 2003, 12:46:48 ; Search time 13.5 Seconds
(without alignments)

2606.653 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 1096

Sequence: 1 tgggaatagttctgttctta.....gggttagtcagattgttg 598

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09914191/runat_09072003_112350_12911/app_query.fasta_1.775
-DB=Issued Patents AA -GPM=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914191 @CIGN 1.1 28 @runat_09072003_112350_12911 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	6.9	419	4	US-09-461-697-389
2	76	6.9	488	4	US-09-461-697-375
3	76	6.9	490	4	US-09-461-697-373
4	76	6.9	490	4	US-09-056-285A-10
5	76	6.9	1167	4	US-09-008-097-6
6	75.5	6.8	393	1	US-08-429-742-2
7	72.5	6.6	453	2	US-08-484-126-3
8	72.5	6.6	1012	4	US-08-811-481-16
9	72	6.6	375	4	US-09-461-697-395
10	71	6.4	247	2	US-08-463-311-2
11	71	6.4	247	4	US-09-140-804-8
12	71	6.4	247	4	US-09-118-408-3

ALIGNMENTS

RESULT 1

US-09-461-697-389
; Sequence 389, Application US/09461697
; Patent No. 6277974

GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.

; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn

; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.

; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.

; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH

; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697

; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389

; LENGTH: 419
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-461-697-389

Alignment Scores:

Pred. No.: 1.71

Score: 76.00

Percent Similarity: 33.70%

Best Local Similarity: 26.09%

Query Match: 6.93%

DB: 4

Length: 419

Matches: 24

Conservative: 7

Mismatches: 21

Indels: 40

Gaps: 4

US-09-914-191-1 (1-598) x US-09-461-697-389 (1-419)

```
QY 323 GAAGCATCCCGATGATATCGCAGTCAAGGCACTGGCTTTGTCTC----- 370
Db 39 GluAlaAspArgAspIleMetThrLeuAlaAsnHisTrpAsnCysProValLeuSerSer 58
QY 371 -----GGTCCGGGTCACTGCCATCTTTTTCCTTCCAT 403
Db 59 AspSerAspPheCysIlePheAspLeuLysThrGlyPheCysProLeuAsnSerPheGln 78
QY 404 TTC-----TGTGGCAG 415
Db 79 TrpArgAsnMetAsnThrIleLysGlyThrGlnAsnTyrIleProAlaLysCysPheSer 98
QY 416 CTTAAATTTCTTTGTCACTCATCCAC-----CTTCTGCCATAT----- 457
Db 99 LeuAspAlaPheCysHisHisPheSerAsnMetAsnLysAlaLeuLeuProLeuPheAla 118
QY 458 -----CAACACAGTCCCTTTTCTTATA 478
Db 119 ValLeuCysGlyAsnAspHisValAsnLeuProIle 130

RESULT 2
US-09-461-697-375
; Sequence 375, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 375
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-375

Alignment Scores:
Pred. No.: 1.82 Length: 488
Score: 76.00 Matches: 24
Percent Similarity: 33.70% Conservative: 7
Best Local Similarity: 26.09% Mismatches: 21
Query Match: 6.93% Indels: 40
DB: 4 Gaps: 4

US-09-914-191-1 (1-598) x US-09-461-697-375 (1-488)

QY 323 GAAGCATCCCGATGATATCGCAGTCAAGGCACTGGCTTTGTCTC----- 370
Db 108 GluAlaAspArgAspIleMetThrLeuAlaAsnHisTrpAsnCysProValLeuSerSer 127
QY 371 -----GGTCCGGGTCACTGCCATCTTTTTCCTTCCAT 403
Db 128 AspSerAspPheCysIlePheAspLeuLysThrGlyPheCysProLeuAsnSerPheGln 147
QY 404 TTC-----TGTGGCAG 415
Db 148 TrpArgAsnMetAsnThrIleLysGlyThrGlnAsnTyrIleProAlaLysCysPheSer 167
QY 416 CTTAAATTTCTTTGTCACTCATCCAC-----CTTCTGCCATAT----- 457
Db 168 LeuAspAlaPheCysHisHisPheSerAsnMetAsnLysAlaLeuLeuProLeuPheAla 187
QY 458 -----CAACACAGTCCCTTTTCTTATA 478
```

```
Db 188 ValLeuCysGlyAsnAspHisValAsnLeuProIle 199

RESULT 3
US-09-461-697-373
; Sequence 373, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 373
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-373

Alignment Scores:
Pred. No.: 1.82 Length: 490
Score: 76.00 Matches: 24
Percent Similarity: 33.70% Conservative: 7
Best Local Similarity: 26.09% Mismatches: 21
Query Match: 6.93% Indels: 40
DB: 4 Gaps: 4

US-09-914-191-1 (1-598) x US-09-461-697-373 (1-490)

QY 323 GAAGCATCCCGATGATATCGCAGTCAAGGCACTGGCTTTGTCTC----- 370
Db 110 GluAlaAspArgAspIleMetThrLeuAlaAsnHisTrpAsnCysProValLeuSerSer 129
QY 371 -----GGTCCGGGTCACTGCCATCTTTTTCCTTCCAT 403
Db 130 AspSerAspPheCysIlePheAspLeuLysThrGlyPheCysProLeuAsnSerPheGln 149
QY 404 TTC-----TGTGGCAG 415
Db 150 TrpArgAsnMetAsnThrIleLysGlyThrGlnAsnTyrIleProAlaLysCysPheSer 169
QY 416 CTTAAATTTCTTTGTCACTCATCCAC-----CTTCTGCCATAT----- 457
Db 170 LeuAspAlaPheCysHisHisPheSerAsnMetAsnLysAlaLeuLeuProLeuPheAla 189
QY 458 -----CAACACAGTCCCTTTTCTTATA 478
Db 190 ValLeuCysGlyAsnAspHisValAsnLeuProIle 201

RESULT 4
US-09-056-285A-10
; Sequence 10, Application US/09056285A
; Patent No. 6403307
; GENERAL INFORMATION:
; APPLICANT: Stone, Edwin M.
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Fingert, John
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
```

Qy	132	TCCCTTCCTCCTCGTGTCCCTCCCAACTTGAG-----	164
Db	718	SerCysGlySerLeuPheProGlyAlaLeuGlnAArgLeuSerArgSerIleValAlaGser	737
Qy	165	TTGTGTCATTCCGCACCACTGTCTGGGTGTAGGGATGCTACAGCCACTTAAGG----	218
Db	738	ArgAlaHisSerThrAlaVal-----GlvilePheSerValLeuLeuValPhe	753

Qy	219	CAAGGAGCCCTGGGAGGTGGGAGGCGCTGCATGGTTAAGCACACC-----AGAACT	269
		:	
Db	754	ThrSerAlaIleAlaAsnMetPheThrCys-----AsnHisThrProIleArgSerCys	771
Qy	270	GAAGCCCAAAAGGTCAGCTGCTTCATCAGAACTCTGTGATCTCTCCAGAAAGCA	329
		:	
Db	772	AlaAlaArgMetLeuAsnLeuThrProAlaAspIleThrAlaCysHisLeuGlnLeu	791
Qy	330	TCCCCGATGATATCG-----CAGTCAAGGCGACTGGCTTTGTCTGTGTCGG	377
		:::	
Db	792	AsnTySerLeuGlyLeuAspAlaProLeuCysGluGlyThr-----	805
Qy	378	GTCACTGCCATCTTTTTCCTTCCATTCTGTGCGCAGCTTAATTTCTTTGTCACT	437
Db	806	MetProThrCysSerPheProGluValSerIleGlyAsnMetLeuSerLeuLeuAla	825
Qy	438	TCATCCACCTTCTGCCCATATCAACACATC	467
Db	826	SerSerValPheLeuHisIleSerSerIle	835

RESULT 6

```

US-08-429-742-2
; Sequence 2, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,742
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-429-742-2

```

Alignment Scores:	1.91	Length:	393
Pred. No.:	Score:	Matches:	34
	75.50	Conservative:	16
Percent Similarity:	37.04%	Mismatches:	46
Best Local Similarity:	25.19%	Indels:	39
Query Match:	6.83%	Gaps:	7
DB:	1		

US-09-914-191-1 (1-598) x US-08-429-742-2 (1-393)

QY 592 TCTGGACTAACA-----CCAGTACAAGCCC AAGGGAGATTACAGGGA 551

```

Db      54 SerGlyPheThrIlePheLeuAsnGluTyrProAlaLeuLysAsnSerLysTyrGln--- 72
Qy     550 ACTCTTNCAGCAGACAAGATGAGATTTTGTTTCTGAATCAACATCAACTATAATAAT 491
Db      73 LeuLeuHisHisSerAlaAsnGlnLeuSerIleThrValProAsnValThrLeuGlnAsp 92
Qy     490 GAGCTGCCGATGTATAGAAAGGACCTGTGTGTATATATGCGCAGAGGTGGATGAAGTGATG 431
Db      93 Glu-----GlyValTyrLysCysLeuHisTyrSerAspSerValSer 106
Qy     430 ACAAAAGAATTAACTGCCAACAGAAATGGAAGGAAAAAGATGGCAGTGCACCGGACC 371
Db     107 ThrLysGluVal-----LysValIleValLeuAlaThr 117
Qy     370 AGGACAAGCCAGTGCCTTGCACCTGCGATATATCFCGGGGATGCTTCTCGGAAGGAACAT 311
Db     118 ProPheLysProfile---LeuGluAlaSerValIleArgLysGlnAsnGlyGluGluHis 136
Qy     310 CCAGAGATTCTAGATGAAGACAGCTGACCCCTTTTGCAGCTTCAGTTCTGTGTGCTTTAAC 251
Db     137 -----ValValLeuMetCysSerThr 143

RESULT 7
US-08-484-126-3
; Sequence 3, Application US/08484126
; Patent No. 5985655
; GENERAL INFORMATION:
; APPLICANT: Anderson, W. French
; APPLICANT: Baltrucki, Leon F.
; APPLICANT: Mason, James M.
; TITLE OF INVENTION: Targetable Vector Particles
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER-READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,126
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,347
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: 08/973,307
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lillie, Raymond J.
; REGISTRATION NUMBER: 31,778
; REFERENCE/DOCKET NUMBER: 271010-281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```



```
;
; FEATURE:
; NAME/KEY: xenotropic gp70 protein
; US-08-484-126-3

Alignment Scores:
Pred. No.: 4,6 Length: 453
Score: 72.50 Matches: 44
Percent Similarity: 33.85% Conservative: 21
Best Local Similarity: 22.92% Mismatches: 59
Query Match: 6.61% Indels: 68
DB: 2 Gaps: 11

US-09-914-191-1 (1-598) x US-08-484-126-3 (1-453)
Qy 113 GATACACACCCCATCC-----ATCCCTTCTTCCCTGTTCCCTC 154
Db 255 AspGlnLeuProSerGlnProValGlnLeuProAArgProHisProPro 274
Qy 155 CCACTTGAGTTGTGTCATTCGACCAAGTGTCTGGTGTGGTAGGATGCTACAGCCACT 214
Db 275 ProSerGlyThrValSerMetValProGlyAla-----ProPro 287
Qy 215 AAGCAGAGGAGCCCTGGAGGTGGAGGGCTTGCATGTTAAGCACACCACTGAAGC 274
Db 288 ProSerGlnProGlyThrGlyAspArgLeuLeuAsn----- 300
Qy 275 GCAAAAGGTCAGCTGCTTCTCATCTAGATCTCTGGATGTTCTTCAGAAAGCATCCC 334
Db 301 -----LeuValGluGlyAlaTyrglnAlaLeuAsnLeuThrSerPro 314
Qy 335 GATGATATCGAGTGCAGGACACTGCTTGT----- 367
Db 315 Asp-----LysThrGlnGluCysTrpLeuCysLeuValSerGlyProProTyrrGlu 332
Qy 368 -----CCTGGTCCGGTCACTGCCATCTT 391
Db 333 GlyValAlaValLeuGlyThrTyrrSerAsnHisThrSerAlaProAlaAsnCysSerVal 352
Qy 392 TTTTCTTCCATTTCTGTGGACGCTTAATTTCTTTTGTCTCATCTTCACTCCACTTCTG 451
Db 353 AlaSerGlnHis-----LysLeuThrLeu-SerGluValThrGlyGlnGlyLeuCy 369
Qy 452 CCATATCAACACAGCTCCCTTCCATATACCGGAGCTCATATTATAGTTGATGTGAA 511
Db 369 s---ValGlyAlaValProLysThrHisGlnAla-----LeuCysAs 382
Qy 512 TTCAGAAAACAATCTCA-----TTCTTGTCTGCTGNAGAGTTTCCCTGTAA 559
Db 382 nThrThrGlnLysThrSerAspGlySerTyrrTyrrLeuAlaAla-----ProAlaG 399
Qy 560 TCTCCCTTGGCTGTGT---ACTGGTGTAGTCCA 590
Db 399 yThrLysThrAlaCysAsnThrGlyLeuThrPro 410

RESULT 8
US-08-811-481-16
; Sequence 16, Application US/08811481
; Patent No. 6300093
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1012 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-811-481-16

Alignment Scores:
Pred. No.: 6,39 Length: 1012
Score: 72.50 Matches: 26
Percent Similarity: 37.37% Conservative: 11
Best Local Similarity: 26.26% Mismatches: 50
Query Match: 6.61% Indels: 12
DB: 4 Gaps: 3

US-09-914-191-1 (1-598) x US-08-811-481-16 (1-1012)
Qy 129 CCATCCCTTCCCTTCCCTGTTCCCTCCCACTTGAGTTGTGTCATTCGCACCACTGCTCT 188
Db 18 ProAlaAlaProSerSerValPro-----HisGlyArgGlnLeuPro 31
Qy 189 GGGTGTAGGATGCTACAGCCACCTAAGCGAAGAGCCCTGGAGGCTGGAGGGGTTGC 248
Db 32 GlyArgLeuGlyCysLeu-----LeuGluGluGlyLeuCysGlyAlaSerGluAlaCys 49
Qy 249 ATGGTTTAAAGCACACCACTGAAGCGCAAAAG-----GGTCAGCTGTCTTCATCT 299
Db 50 ValAsnAspGlyValPheGlyArgCysGlnLysValProAlaMetAspPheTyrrArgTyr 69
Qy 300 AGAATCTCTGGATGTTCTTCAGAAAGCATCCCGATGATATCGCAGTGCAGAGGCACT 359
Db 70 GluValSerProValAlaLeuGlnArgLeuArgValAlaLeuGlnLysLeuSerGlyThr 89
Qy 360 GGTCTTGTCTGTCGGGTCACTGCATCTTTTTCCTTCCATTTCTCTGTGCA 414
Db 90 GlyPhe-ThrTrpGlnAspTyrrThrGlnTyrrValMetAspGlnGluLeuAla 107

RESULT 9
US-09-461-697-395
; Sequence 395, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
```

```
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-395

Alignment Scores:
Pred. No.: 4.99 Length: 375
Score: 72.00 Matches: 375
Percent Similarity: 32.56% Conservative: 22
Best Local Similarity: 25.58% Mismatches: 6
Query Match: 6.57% Indels: 18
DB: 40 Gaps: 4

US-09-914-191-1 (1-598) x US-09-461-697-395 (1-375)
QY 341 ATCGCAGTCAAGGCGCACTGGCTTGTCTT----- 370
Db 1 MetThrLeuAlaAsnHisTrpAsnCysProValLeuSerSerAspSerAspPheCysIle 20
QY 371 -----GGTCGGGTCACTGCATCTTTTTCCTTCCATTTC----- 406
Db 21 PheAspLeuIysThrGlyPheCysProLeuAsnSerPheGlnTrpArgAsnMetAsnThr 40
QY 407 -----TGTTGGCAGCTTAATTTCTTTTCTTCAT 433
Db 41 IleLysGlyThrGlnAsnTyrlleProAlaLysCysPheSerLeuAspAlaPheCysHis 60
QY 434 CACTTCATCCAC-----CTTCTGCCATAT-----CAA 460
Db 61 HisPheSerAsnMetAsnLysAlaLeuLeuProLeuPheAlaValLeuCysGlyAsnAsp 80
QY 461 CACAGTCCCTTCTCTATA 478
Db 81 HisValAsnLeuProIle 86

RESULT 10
US-08-463-911-2
; Sequence 2, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-463-911-2

Alignment Scores:
Pred. No.: 5.42 Length: 247
Score: 71.00 Matches: 15
Percent Similarity: 51.28% Conservative: 5
Best Local Similarity: 38.46% Mismatches: 11
Query Match: 6.43% Indels: 8
DB: 2 Gaps: 2

US-09-914-191-1 (1-598) x US-08-463-911-2 (1-247)
QY 241 CTCCACCTCCAGGGCTCCTTGGCTTAGTGGCTTAGCAT---CCCTACCACCCAGGA 185
Db 32 ValProProProlysGlyThrCysAlaGlyTrpMetAlaGlyIleProGlyHisProGly 51
QY 184 CACTGCTGCGAATGACACAACTCAAGTTGGAGGGGAACAGGGAAGAGGGATGGA 128
Db 52 His-----AsnGlyThrProGlyArgAspGlyArgAspGly 63

RESULT 11
US-09-140-804-8
; Sequence 8, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-8

Alignment Scores:
Pred. No.: 5.42 Length: 247
Score: 71.00 Matches: 15
Percent Similarity: 51.28% Conservative: 5
Best Local Similarity: 38.46% Mismatches: 11
Query Match: 6.43% Indels: 8
DB: 2 Gaps: 2

US-09-914-191-1 (1-598) x US-09-140-804-8 (1-247)
QY 241 CTCCACCTCCAGGGCTCCTTGGCTTAGTGGCTTAGCAT---CCCTACCACCCAGGA 185
Db 32 ValProProProlysGlyThrCysAlaGlyTrpMetAlaGlyIleProGlyHisProGly 51
QY 184 CACTGCTGCGAATGACACAACTCAAGTTGGAGGGGAACAGGGAAGAGGGATGGA 128
Db 52 His-----AsnGlyThrProGlyArgAspGlyArgAspGly 63

RESULT 12
US-09-118-408-3
; Sequence 3, Application US/09118408A
; Patent No. 6265544
; GENERAL INFORMATION:
```



```
; Sequence 2, Application US/08952981A
; Patent No. 6383768
; GENERAL INFORMATION:
; APPLICANT: DE FRANCESCO, Raffaele
; APPLICANT: TOMEI, Licia
; APPLICANT: BEHRENS, Sven-Erik
; TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE RNA-DEPENDENT RNA
; TITLE OF INVENTION: POLYMERASE AND TERMINAL NUCLEOTIDYL TRANSFERASE
; TITLE OF INVENTION: ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)
; FILE REFERENCE: IT0002P
; CURRENT APPLICATION NUMBER: US/08/952,981A
; CURRENT FILING DATE: 1998-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: cdna clone pCD (38-9.4)
US-08-952-981A-2
```

```
Alignment Scores:
Pred. No.: 15.2 Length: 2201
Score: 70.50 Matches: 24
Percent Similarity: 41.57% Conservative: 13
Best Local Similarity: 26.97% Mismatches: 33
Query Match: 6.38% Indels: 19
DB: 4 Gaps: 4
```

```
US-09-914-191-1 (1-598) x US-08-952-981A-2 (1-2201)
QY 450 AGAAGTGGATGAAGTCATGACAAAGAAATTA----- 418
Db 1183 LysThrTrpLeu-----GlnSerLysLeuLeuProGlnLeuProGlyValProPhe 1199
QY 417 ---AGTCGCCACAGAAATGAAGAAAGAAAGATGCGAGTACCCGGACCCAGGACAAAGC 361
Db 1200 PheSerCysGlnArgGlyTyLysGlyVal---TipArgGlyAspGlyIleMetGlnThr 1218
QY 360 CAGTGCCTTGCCTGCGATATCATCGGGATGCTTCTCGAAGGAACATCCAGAGATTC 301
Db 1219 ThrCysProCysGlyAlaGlnIleThrGlyHisValLysAsnGlySerMetArgIleVal 1238
QY 300 TAGATGAAGACAGCTGAC-----CCTTTTCGCTTCAGTTCTCGTGTGCT 256
Db 1239 GlyProLysThrCysSerAsnThrTrpHisGlyThrPheProIleAsnAlaTyThrThr 1258
QY 255 TAACCATGCAAGCCCTCCACCTCCCA 229
Db 1259 GlyProCysThrProSerProAlaPro 1267
```

Search completed: July 9, 2003, 12:51:58
Job time : 19.5 secs

Result No.	Score	Query Match	Length	DB	ID	Description
C	1	455	41.2	85	9	US-10-106-698-5245
	2	88.5	8.1	457	9	US-09-934-455-302
	3	80	7.3	233	10	US-09-925-300-1529
	4	78	7.1	833	10	US-09-844-281-1
	5	78	7.1	833	10	US-09-844-281-1
C	1	455	41.2	85	9	US-10-106-698-5245
	2	88.5	8.1	457	9	US-09-934-455-302
	3	80	7.3	233	10	US-09-925-300-1529
	4	78	7.1	833	10	US-09-844-281-1
	5	78	7.1	833	10	US-09-844-281-1

```

Db      267 ValProLeuAsnGlnThrAsnGluGluAlaGlyThrVal-----SerProLeuProLys 284
QY     131 ATCCCTTCTCCTCCCTGGTTCCCTCCCAACTTCAGTTGTGTCATTCGCACACAGTGTCCTCG 190
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      285 ValProCysPheProGlyProProProThrTrpProTyraLaTrpAsnGlyValSerTrp 304
QY     191 GTGGTAGGAGTCTCACGCCACCTAAGGCAAG- - - - - AGCCCTGGG 232
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      305 ThrIleLeuProPheTyProProProAlaTyTrpSerCysProGlyValSerProGly 324
QY     233 AGTGGGAGGGCTTCATGGTTTAAGCACACACAGAAGTGAAGCGCAAAGGTCAGCTGTC 292
       |||||
Db      325 AlaTrpAsnSerPhe----- 329
QY     293 TTCATCTAGATCTCTCGATGTTCTTCACAAAGC 328
       |||||
Db      330 -----ThrTrpMetProGlnProAsnSer 337

RESULT 3
US-09-925-300-1529
; Sequence 1529, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1529
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1529

Alignment Scores:
Pred. No.:          7.12           Length:         233
Score:             80.00          Matches:        27
Percent Similarity: 40.40%        Conservative:   13
Best Local Similarity: 27.27%    Mismatches:    26
Query Match:       7.30%         Indels:        33
DB:                10            Gaps:           2

US-09-914-191-1 (1-598) x US-09-925-300-1529 (1-233)
QY     87 CAAGGTGTTCTGCTTAAACAGAGTAAGATACACCACCCCCTCATCCCTCTCTCCCTG 146
       ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      85 GlnGlySerLeuValAsnGlnAsnLeuLeuHisGlnHisGlnThr----- 100
QY     147 TTCCCCCTCCCAACTTGAGTTGTGTATTTCGCACACAGTGTCTCTGGGTGTAGGATGCTAC 206
       |||||
Db      100 ----- 100
QY     207 AGCCACCTAAGCAAGGAGCCCTGGGAGGTGGAGGGCTTGCATGGTTAAGCACACCAGA 266
       ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      101 -----GlnGlyAlaLeuGlyGlySerArgAlaLeu-----SerAsn 112
QY     267 ACTGAAGCGCAAAGGGTCAGCTGCTTCATCTAGAACTCTGGATGTTCTTCCACAGAAA 326
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      113 SerValSerAsnMetGlyLeuSerGluSerSerLeuGlySerAlaLysHisGlnGln 132
QY     327 GCATCCCCGATGATATCGCAGTGAAGGGCACTGGCTTTGTCTGGTCCGGGTCA 381
       ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      133 GlnSerProVal-SerGlnSerMetGlnThrLeuSerAspSerLeuSerGlySer 150

RESULT 4
US-09-844-281-1

```


US-09-922-261-373

Alignment Scores:
Pred. No.: 23.5 Length: 490
Score: 76.00 Matches: 24
Percent Similarity: 33.70% Conservative: 7
Best Local Similarity: 26.09% Mismatches: 21
Query Match: 6.93% Indels: 40
DB: 10 Gaps: 4

US-09-914-191-1 (1-598) x US-09-922-261-373 (1-490)

QY 323 GAAGCATCCCGATGATCGAGTGAAGGCAAGGCACTGGTGTGCTCT----- 370
Db 110 GluAlaAspArgAspIleMetThrLeuAlaAsnHisTrpAsnCysProValLeuSerSer 129
QY 371 -----GGTCCGGTCACTGCCATCTTTTCTCTCCAT 403
Db 130 AspSerAspPheCysIlePheAspLeuLysThrGlyPheCysProLeuAsnSerPheGln 149
QY 404 TTC-----TGTGGCAG 415
Db 150 TrpArgAsnMetAsnThrIleLysGlyThrGlnAsnTyrlleProAlaLysCysPheSer 169
QY 416 CTTAATTCTTTGTCTATCTCTCATCCAC-----CTTGTGCCATAT----- 457
Db 170 LeuAspAlaPheCysHisHisPheSerAsnMetAsnLysAlaLeuLeuProLeuPheAla 189
QY 458 -----CAACACAGTCCCTTTCCTATA 478
Db 190 ValLeuCysGlyAsnAspHisValAsnLeuProIle 201

RESULT 9

US-09-750-240-6
; Sequence 6, Application US/09750240
; Patent No. US20020103147A1
; GENERAL INFORMATION:
; APPLICANT: Hammon, H. K.
; APPLICANT: Insel, P. A.
; APPLICANT: Ping, P.
; APPLICANT: Post, S. R.
; APPLICANT: Gao, M.
; TITLE OF INVENTION: GENE THERAPY FOR CONGESTIVE HEART
; FILE REFERENCE: FAILURE
; CURRENT APPLICATION NUMBER: US/09750,240
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 09/472,667
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: US 09/008,097
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: US 08/924,757
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/048,933
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: US 08/708,661
; PRIOR FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-750-240-6

Alignment Scores:
Pred. No.: 30.1 Length: 1167
Score: 76.00 Matches: 32
Percent Similarity: 41.54% Conservative: 22
Best Local Similarity: 24.62% Mismatches: 46
Query Match: 6.93% Indels: 30
DB: 10 Gaps: 7

US-09-914-191-1 (1-598) x US-09-750-240-6 (1-1167)

QY 132 TCCCTTCCTCCTCTGTTCCCTCCCAACTTGAG----- 164
Db 718 SerCysGlySerLeuPheProLysAlaLeuGlnArgLeuSerArgSerIleValArgSer 737
QY 165 TTGTGTTCATCCAGCCAGTGTCTCTGGGTGGTAGGATGTACAGCCACCTAAGG----- 218
Db 738 ArgAlaHisSerThrAlaVal-----GlyIlePheSerValLeuLeuValPhe 753
QY 219 CRAAGGAGCCCTGGGAGGTGGAGGGCTTCATGGTTAAGCACACC-----AGAACT 269
Db 754 ThrSerAlaIleAlaAsnMetPheThrCys-----AsnHisThrProIleArgSerCys 771
QY 270 GAAGCCGAAAAGGTCAGTCTTCATCTAGATCTCTGGATGTCTCTCCAGAAAGCA 329
Db 772 AlaAlaArgMetLeuAsnLeuThrProAlaAspIleThrAlaCysHisLeuGlnLeu 791
QY 330 TCCCGCATGATATCG-----CAGTCAAGGCGACCTGGCTTGTCTGTGTCGG 377
Db 792 AsnTySerLeuGlyLeuAspAlaProLeuCysGluGlyThr----- 805
QY 378 GTCACTGCCCATCTTTTCTCTCCATTTCTGTGGCAGCTTAATTTCTTTTGTCACTACT 437
Db 806 MetProThrCysSerPheProGluValSerIleGlyAsnMetLeuSerLeuLeuAla 825
QY 438 TCATCCACCTTCTGCCATATCAACACAGTC 467
Db 826 SerSerValPheLeuHisIleSerSerIle 835

RESULT 10

US-09-984-130-144
; Sequence 144, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 144
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-144

Alignment Scores:
Pred. No.: 25 Length: 393
Score: 75.50 Matches: 34
Percent Similarity: 37.04% Conservative: 16
Best Local Similarity: 25.19% Mismatches: 46
Query Match: 6.83% Indels: 39
DB: 9 Gaps: 7

US-09-914-191-1 (1-598) x US-09-984-130-144 (1-393)

QY 592 TCTGGACTAACA-----CCAGTACAAGCCCAAGGAGATTACAGGA 551
Db 54 SerGlyPheThrIlePheLeuAsnGluTyProAlaLeuLysAsnSerLysTyGln--- 72
QY 550 ACTCTTNCAGCAGACAAGAAATGAGATTTTGTCTTCTCAATCAACTATAATAAT 491

```
Db 73 LeuLeuHisHisSerAlaAenGlnLeuSerIleThrValProAenValThrLeuGlnAsp 92
QY 490 GAGCTCCGATGTATAGGAAGGACTGCTTTATATATGCGAGAGGTGGATGAGTGATG 431
Db 93 Glu-----GlyValTyrLysCysLeuHisTyrSerAspSerValser 106
QY 430 ACAAAGAAATTAAGCTGCCAACAGAAATGGAAGGAAAAAGATGGCAGTGGCCCGGACC 371
Db 107 ThrLysGluVal-----LysValIleValLeuAlaThr 117
QY 370 AGGACAAAGCAGTCCCTTGCACATGCGATATCATCGGGATGCTTTCTGGAAGGAACAT 311
Db 118 ProPheLysProIle---LeuGluAlaSerValIleArgLysGlnAenGlyGluGluHis 136
QY 310 CCAGAGATTCATAGATGAGACAGTGCACCTTTTGGCTTCAGTTCCTGGTGCTCTAAC 251
Db 137 -----ValValLeuMetCysSerThr 143
QY 250 ATCAAGCCCTCCACCTCCAGGCTCCTTGGCTTAGTGGCTG 206
Db 144 MetArgSerLysProProGln-----IleThrTrpLeu 155
```

RESULT 11

```
US-09-925-300-1578
; Sequence 1578, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1578
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (209)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1578
```

```
Alignment Scores:
Pred. No.: 28.3 Length: 393
Score: 75.00 Matches: 29
Percent Similarity: 39.47% Conservative: 16
Best Local Similarity: 25.44% Mismatches: 45
Query Match: 6.84% Indels: 24
DB: 10 Gaps: 7
```

US-09-914-191-1 (1-598) x US-09-925-300-1578 (1-393)

```
QY 58 CCAAGGAACCATGTTCCACACCGCAACCAAGGTGTTCTGTCTAAACAGATGAAGATAC 117
Db 292 ProArgArgProCysArgSerIleLysGlyAlaTrpLeuCysCysLeuGlnSerArgThr 311
QY 118 ACCACCCCATCATCC-----TTCTTCCCTGTTCCCTCCCACTTGAGT 165
Db 312 ThrArgLeuThrArgProSerThrThrThrPheLeuSerCysSerProPro-----Ala 329
QY 166 TGTGTATTGCGACCATGTCCTGGGTGGTAGGATGCTACACCCACTTAAGCAAGGAG 225
Db 330 AlaProLeuHisProSerThrMetGlySerArgSerProProLeuGln---GlyArgAla 348
QY 226 CCCTGGGAGGTGGGAGGGCTTGCATGTTAAGCACACCAAGCAAGTGAAGCGCAAAAGGGTC 285
```

```
Db 349 ProGlnPro-----ArgSerTrpThrGlyThrArgArgArgLysArgArgMet 365
QY 286 AGCTGCTTCATCTAGAATCTCTGGATGTTCCTCCAGAAAGCATCCCGATGATATCGC 345
Db 366 -----MetLysMetGluLysMet-----ArgArgLysSerPro----- 376
QY 346 AGTCAAGGCGACTGCTTTGCTTGGTCCGGGTCACTGCCA 387
Db 377 -----SerAlaSerGlyLysGlySerArgPro 385
```

RESULT 12

```
US-10-184-644-107
; Sequence 107, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 107
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-107
```

```
Alignment Scores:
Pred. No.: 69 Length: 2478
Score: 73.50 Matches: 29
Percent Similarity: 35.05% Conservative: 5
Best Local Similarity: 29.90% Mismatches: 48
Query Match: 6.65% Indels: 15
DB: 9 Gaps: 4
```

US-09-914-191-1 (1-598) x US-10-184-644-107 (1-2478)

```
QY 357 TGCCTTGCACCTGCGATATCATCGGGATGCTTTCTGGNAGGAACATCCAGAGATTCTAG 298
Db 1125 CysThrCysAlaAlaGlyThrAlaGlyCysThrGlyGly----- 1137
QY 297 ATGAAGACAGCTGACCCCTTTTCGGCTTCAGTTCTGGTGTCTTAACCATGCAAGCCCTCC 238
Db 1138 ---GlyAlaCysThrAlaCysAlaGlyGlyThrGlyCysAlaThrGlyCysCysAlaCys 1156
QY 237 CACCTCCAGGCTCCTTGGCTTAGTGGCTGTAGCATCCCTACCAAGGAGGAGGAGTGGT 178
Db 1157 CysAlaThrGlyCysCysAla---GlyGlyCysThrAlaAlaThrThrThrThrGly 1175
QY 177 GCGAATGACACAACTCAAGTTGGGAGGGAACA---GGGAAGGAGGAGGATGGATGGGGT 121
Db 1176 GlyThrGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1195
QY 120 GGTGTATCTTACTCTGTTTAAAGCAGAACACCTTGTTCGGGTGTTTGGGAACA 70
Db 1196 GlyThrThrThrThrGly-----CysCysAlaThrGlyThr 1207
```

RESULT 13

```
US-10-184-634-107
; Sequence 107, Application US/10184634
```

```
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 107
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-107

Alignment Scores:
Pred. No.: 69          Length: 2478
Score: 73.50          Matches: 29
Percent Similarity: 35.05% Conservative: 5
Best Local Similarity: 29.90% Mismatches: 48
Query Match: 6.65%     Indels: 15
DB: 9                 Gaps: 4

US-09-914-191-1 (1-598) x US-10-184-634-107 (1-2478)
QY 357 TGCCTTGCACTGGGATATCATCGGGATGCTTCTGGAAGGAACATCCAGATTCTAG 298
Db 1125 CysThrCysAlaAlaGlyThrAlaGlyCysThrGlyGly-----1137
QY 297 ATGAAGACACGTGACCCCTTTTGGGCTTCAGTCTGGTGCTTAACCATGCAAGCCCTCC 238
Db 1138 ---GlyAlaCysThrAlaCysAlaGlyGlyThrGlyCysAlaThrGlyCysAlaCys 1156
QY 237 CACCTCCCAAGGCTCCTTGCTTAGTGCTGAGTACCTCCACCCAGGACACTGGT 178
Db 1157 CysAlaThrGlyCysCysAla---GlyGlyCysThrAlaAlaThrThrThrGly 1175
QY 177 GCGAATGACACAACCTCAAGTTGGAGGGGAACA---GGGAAGGAAGGATGGATGGGGT 121
Db 1176 GlyThrGlyThrThrThrThrThrThrThrThrAlaGlyAlaGlyAlaCysThrGlyGly 1195
QY 120 GGTGTATCTTACTCTGTTTAAGCAGAACACCTTGTGCGGTGTTGGAACA 70
Db 1196 GlyThrThrThrThrThrGly-----CysCysAlaThrGlyThr 1207

RESULT 14
US-09-896-738-18
; Sequence 18, Application US/09896738
; Patent No. US20020165347A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Michael
; APPLICANT: Sullivan, John K.
; APPLICANT: Fang, Mei
; TITLE OF INVENTION: B7-Like Molecules and Uses Thereof
; FILE REFERENCE: 00-513-A
; CURRENT APPLICATION NUMBER: US/09/896,738
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,645
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 18

; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 107
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-107

Alignment Scores:
Pred. No.: 69          Length: 2478
Score: 73.50          Matches: 29
Percent Similarity: 35.05% Conservative: 5
Best Local Similarity: 29.90% Mismatches: 48
Query Match: 6.65%     Indels: 15
DB: 9                 Gaps: 4

US-09-914-191-1 (1-598) x US-09-896-738-18 (1-584)
QY 74 CCAACACCCGCAACAAAGGTGTTCTGCTTAAACAGAGTAAGATACACACCCCATCCATC 133
Db 173 ProGlnProGlnIleIleYsTrpSer-----AspThrLysGlyGluAsnIle 187
QY 134 CTTTCCTTCCTGTTTCCCTCCCAACTTGAGTTGTGTCATTCCGACCACTGTCCTGGTG 193
Db 188 ProAlaValGluAlaProVal-----ValAlaAspGly 198
QY 194 GTAGGATGCTAC-----AGCCACCTAAGGCAAGGAGCCCTGGAGGTGGAG 241
Db 199 ValGlyLeu-TyrAlaValAlaAlaSerValIleMetArgGlySerSerGlyGlyGlyVa 218
QY 242 GGTTCATGTTTAAGCACACACAGAACTGAAGCGCAAAAGGGTCAGCTGCTTCTCTCTAG 301
Db 218 lSerCysIleIleIleArgAsnSerLeuLeuGlyLeuGlyLys-----ThrAlaSerIleSe 236
QY 302 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCATGCAAGGGCACT 359
Db 236 rIleAlaAspProPheArgSerAlaGlnProTrpIleAlaAlaLeuAlaGlyThr 255

RESULT 15
US-09-910-174A-16
; Sequence 16, Application US/09910174A
; Patent No. US20020106730A1
; GENERAL INFORMATION:
; APPLICANT: Coyle, Anthony J.
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: B7-H2 Molecules, No. US20020106730A1el Members of the B7
; FILE OF INVENTION: Family and Uses Thereof
; FILE REFERENCE: 35800/236924
; CURRENT APPLICATION NUMBER: US/09/910,174A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/620,461
; PRIOR FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-910-174A-16

Alignment Scores:
Pred. No.: 51.8        Length: 584
Score: 73.00          Matches: 28
Percent Similarity: 48.00% Conservative: 20
Best Local Similarity: 28.00% Mismatches: 31
Query Match: 6.66%     Indels: 21
DB: 10               Gaps: 4

US-09-914-191-1 (1-598) x US-09-910-174A-16 (1-584)
QY 74 CCAACACCCGCAACAAAGGTGTTCTGCTTAAACAGAGTAAGATACACACCCCATCCATC 133
Db 173 ProGlnProGlnIleIleYsTrpSer-----AspThrLysGlyGluAsnIle 187
QY 134 CTTTCCTTCCTGTTTCCCTCCCAACTTGAGTTGTGTCATTCCGACCACTGTCCTGGTG 193
```

```
Db      |||::: |||::: |||:::
188 ProAlaValGluAlaProVal-----ValAlaAspGly 198
Qy      |||::: |||::: |||:::
194 GTAGGGATGCTAC-----AGCCACCTAAGCGAAGAGCCCTGGAGTGGAG 241
Db      |||::: ||| |||::: |||::: |||::: |||:::
199 ValGlyLeu-TyrAlaValAlaAlaSerValIleMetArgGlySerSerGlyGlyVa 218
Qy      |||::: |||::: |||::: |||::: |||::: |||:::
242 GGCTTGCATGGTTAAGCACACCACTGAAGCGCAAAGGTCAGTGTCTTCATCTAG 301
Db      |||::: |||::: |||::: |||::: |||::: |||:::
218 lSerCysIleIleArgAsnSerLeuLeuGlyLeuGluLys-----ThrAlaSerIleSe 236
Qy      |||::: |||::: |||::: |||::: |||::: |||:::
302 AATCTCTGGATGTTCTTCAGAAAGCATCCCGATGATATCGACGTGCAAGGGCACT 359
Db      |||::: |||::: |||::: |||::: |||::: |||:::
236 rIleAlaAspProPhePheArgSerIleGlnProTrrIleAlaAlaLeuAlaGlyThr 255
```

Search completed: July 9, 2003, 12:53:09
Job time : 39 secs

QY 475 AGGAAGGAGCTGTTGATATGCGAAGGTGCGATGAAGTATGACAAAGAAATTAAG 416
 Db LysLysGlySerIleTrpIleArgGluProIleAspGlnGluTrpHisGlnGlnAspLys 225
 QY 415 CTGCAACAGAAATGGAAGAAAGAAAGATGCGAGTACCCGGACGAGCAAAAGCCAGTG 356
 Db LysPheSerValLysGlnLysLysLysMetValLeuSer----- 238
 QY 355 CCCTTGACATGCGATATCATCGGGGATGCTTTCTGGAAGGAACATCCA-----GAGATT 302
 Db LysLysHisValSerLeuIleAspAspPheTrpThrSerArgProPheLeuGluVal 258
 QY 301 CTA 299
 Db Lys 259 Leu 259

RESULT 2
 S64315
 hypothetical protein YGR024c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G4041
 C:Species: Saccharomyces cerevisiae
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
 C:Accession: S64315
 R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64071
 A:Accession: S64315
 A:Molecule type: DNA
 A:Residues: 1-237 <RIE>
 A:Cross-references: EMBL:272809; NID:g1322995; PID:e243373; PID:g1322996; GSPDB:GN00007;
 A:Experimental source: strain S288C
 C:Genetics:

Alignment Scores:
 Pred. No.: 0.000189 Length: 237
 Score: 122.00 Matches: 23
 Percent Similarity: 75.00% Conservative: 10
 Best Local Similarity: 52.27% Mismatches: 11
 Query Match: 11.04% Indels: 0
 DB: 2 Gaps: 0

US-09-914-191-1 (1-598) x S64315 (1-237)

QY 589 GGACTAACACAGTACAGGCCAAGGAGGATTTACAGGGAATCTTNCAGCAGACAAGAAAT 530
 Db GlyLeuThrProGlnGluSerGluLysLeuCysGlyThrPheSerAsnGluLysGln 191
 QY 529 GAGATTTGTTTCTGAATTCACACTCAACTATAATGAGCTGCGATGATAGGAAA 470
 Db GluIleLeuPheSerGluCysGlyIleAsnTrpAsnAsnGluProGluMetPheLysLys 211
 QY 469 GGGACTGTGTG 458
 Db Lys 212 GlySerLeuVal 215

RESULT 3
 G84731
 hypothetical protein At2g32330 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
 C:Accession: G84731
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.E.; Umayam, L.; Tallon, L.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84731

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-297 <STO>
 A:Cross-references: GB:AE002093; NID:g3831463; PIDN:AAC69945.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g32330
 A:Map position: 2
 C:Superfamily: Saccharomyces cerevisiae hypothetical protein YGR024c

Alignment Scores:
 Pred. No.: 0.182 Length: 297
 Score: 94.00 Matches: 20
 Percent Similarity: 62.75% Conservative: 12
 Best Local Similarity: 39.22% Mismatches: 19
 Query Match: 8.51% Indels: 0
 DB: 2 Gaps: 0

US-09-914-191-1 (1-598) x G84731 (1-297)

QY 595 CAATCTGGACTAACACACAGGAGGAGGATTTACAGGGAATCTTNCAGCAGAC 536
 Db LysSerGlyLysSerLysIleGlnAlaGlnAspTyrLeuLysGlyThrGlnThrArgGlu 246
 QY 535 AAGAATGAGATTTGTTTCTGAATTCACACTCAACTATAATGAGCTGCCGATGAT 476
 Db LysAsnGluLeuLeuSerGlnGlnPheGlyIleGluTyrAsnSerLeuProValIlePhe 266
 QY 475 AGGAAGGAGCTGTGTGATATGCGAAGGAGTG 443
 Db ArgMetGlySerSerValPheArgLeuLysVal 277

RESULT 4

E84722

hypothetical protein At2g31580 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84722

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, L.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84722

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <STO>

A:Cross-references: GB:AE002093; NID:g4582470; PIDN:AAD24854.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g31580

A:Map position: 2

Alignment Scores:

Pred. No.: 0.506 Length: 457

Score: 90.00 Matches: 20

Percent Similarity: 57.14% Conservative: 12

Best Local Similarity: 35.71% Mismatches: 24

Query Match: 8.14% Indels: 0

DB: 2 Gaps: 0

US-09-914-191-1 (1-598) x E84722 (1-457)

QY 595 CAATCTGGACTAACACACAGGAGGAGGATTTACAGGGAATCTTNCAGCAGAC 536
 Db LysSerGlyLysAsnLysThrGlnSerGlnAspTyrLeuLysGlyThrGlnThrArgGlu 413
 QY 535 AAGAATGAGATTTGTTTCTGAATTCACACTCAACTATAATGAGCTGCCGATGAT 476
 Db LysAsnGluLeuLeuSerArgGlnPheGlyIleGluTyrAsnSerLeuProValIlePhe 433

QY 475 AGGAAGGAGCTGTGTGATATGCGAAGGAGTGATGAAGTATGATGACA 428
 Db ArgMetGlySerSerValPheArgLeuLysValSerIleAlaLeuThr 449

Db 69 ArgProGluLysLysGlnSerAspLysSerAsnTyrAlaArgAlaGluLeuPheArgGly 88
 QY 370 TGGTCCGGTCA----- 381
 Db 89 LysSerGlySerValSerPheAsnGlyLeuThrHisGlnLeuValGluSerLysLeu 108
 QY 381 ----- 381
 Db 109 ValSerAlaProPheGlnGluLysGlySerPheLeuTrpValLeuAlaProValVal 128
 QY 382 -----CTCCCATCTTTTCTCCATCTCTCCATCTCTGGCAGCTTAATTT 423
 Db 129 LeuLeuSerSerLeuLeuLeuProGlnPheLeuSerGlyLeuLeuGluAlaThrPhe 148
 QY 424 CTTTGTGCATCTTCATCCACCTTCGCATATCAACA-----CAGTCCCTT 471
 Db 149 LysAsnAspThrValAlaGlyArgSerGluLeuValThrSerPheCysPheGluThrVal 168
 QY 472 TCCTATACATCGGCAGCTCATTTATATAGT 501
 Db 169 PheTyrAlaGlyLeuAlaIlePheLeuSer 178
 RESULT 8
 F84550
 hypothetical protein At2g17310 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: F84550
 R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: F84550
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-370 <SPO>
 A:Cross-references: GB:AE002093; MID:g2262162; PIDN:AAB86508.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g17310
 A:Map position: 2
 C:Superfamily: Arabidopsis thaliana hypothetical protein F24M12.210
 Alignment Scores:
 Pred. No.: 3.95 Length: 370
 Score: 81.50 Matches: 47
 Percent Similarity: 38.12% Conservatives: 30
 Best Local Similarity: 23.27% Mismatches: 68
 Query Match: 7.38% Indels: 57
 DB: 2 Gaps: 10
 US-09-914-191-1 (1-598) x F84550 (1-370)
 QY 520 TTTTCTGAATTCACATCACTATAATAGTGCCTGCGC----- 482
 Db 134 PheCysValPheGlyLeuGlyTyrAspAsnSerArgProGluLysValTyrLysIleLeu 153
 QY 481 -----ATGTATAGGAAGGACTGTGTTGATATGCAGAGGTG-----GATGAAGTG 434
 Db 154 GlyHisLeuPheCysHisGlyLysValLeuArgAspGlnLysValValIleTyrGluCys 173
 QY 433 ATGACAAAGAAATTAACTGCTGCCACAGAAATCGAAGGAAAAAGATGGCAGTGCAGCCGG 374
 Db 174 AlaSerAspSerLeuArgPheIleAspArgProGluAspAspTrpProIleThrGlu 193
 QY 373 ACCAGGACAAAGCAGTGCCTTCGACTGCATATCATCGGGATGCTTCTCG----- 320
 Db 194 ThrAlaLysArgSer-----AsnValSerLeuAsnGlyAsnLeuTyrTrpPheGly 210
 QY 319 -----AAGGAACATCCAGAGATTCAGAT-----GAAGAC 290
 Db 211 CysSerAsnTyrGluAsnAspGluTyrTyrIleArgIlePheAspPheSerThrGluAsp 230

QY 289 AGCTGACCCCTTTTGGCTTCAGTCTGGTGTGCTTAACCATGACGACCCCTCCACCTCCC 230
 Db 231 PheLysProPheCys-----LeuLeu-ProCysGlnMetSerHisSerTh 245
 QY 229 AGGGCTCTCTGGCTTAGGT-----GGCTGTAG 203
 Db 245 rAspGluLeuValLeuAlaValTyrLysGlyAspArgPheSerLeuLeuLysGlnCysSe 265
 QY 202 CATCCCT-----ACCACCAGGACACTGGTGGCAATGACACAACTCA 161
 Db 265 rValThrArgGluIleGlyValTrpValThrLysGluArgIleSerAsnAsp----- 282
 QY 160 AGTTGGGAGGACAGGAGGAGGATGGGTGGGTGGTGTATCTTACTCTCTTTA 101
 Db 283 -----AsnGlyAsnGlyGlyGluValGlu-TrpLeuLysLeuMetThrLeuSerL 300
 QY 100 AGCAGAAC-----ACCTTGTTCGGTGTGGACATGTTCTTCCTTGGCAGAGTCTT 47
 Db 300 ysProAsnLeuProLysLeuPheGlyThrValSerTyrPheIleTyrGlyLysThrLeuT 320
 QY 46 TT 45
 Db 320 yr 320
 RESULT 9
 T51779
 non-phototropic hypocotyl 3-like protein - Arabidopsis thaliana
 N:Alternate names: protein F28D10.10
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
 C:Accession: T51779
 R.; Delzeny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Rudd, S.;
 submitted to the Protein Sequence Database, August 2000
 A:Reference number: Z25454
 A:Accession: T51779
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-661
 A:Cross-references: EMBL:AL391254
 A:Experimental source: cultivar Columbia; BAC clone F28D10
 C:Genetics:
 A:Map position: 3
 A:Introns: 28/3; 83/3; 419/3
 A:Note: F28D10.10
 C:Superfamily: Arabidopsis hypothetical protein F19F18.80
 Alignment Scores:
 Pred. No.: 4.21 Length: 661
 Score: 81.50 Matches: 45
 Percent Similarity: 41.34% Conservatives: 29
 Best Local Similarity: 25.14% Mismatches: 62
 Query Match: 7.38% Indels: 43
 DB: 2 Gaps: 9
 US-09-914-191-1 (1-598) x T51779 (1-661)
 QY 505 ATCACTATAATAATGAGCTCGCATGTATAGAAAGGACTGTG----- 461
 Db 398 LeuAsnTyrSerAspSerGluThrLeuTyrAsnValAspCysValGluArgIleValArg 417
 QY 460 TTGATATGCCACAGGTGATCAAGTGCACAAA-----GAAATTAAGCTG--- 413
 Db 418 HisPheTrpArgLeuValAspSerTyrMetAlaGluAlaSerAspValAsnLeuLys 437
 QY 412 CCAACAGAAATGAAGGAAAAAGATGGCAGTGACCGGACGAGGACAAAGCAGTGCCT 353
 Db 438 ProAspLysMetArgSerLeuAlaAlaLeuProGluSerSerArg-----Pro 454
 QY 352 TTGCTACTGGCATATCATCGG-----GATGCTTCTGGAGCAACATCCAGATCTTA 299
 Db 455 LeuTyrAspGlyLeuTyrArgAlaPheAspIleTyrPheLysGluHisProTrpLeuSer 474

Qy 298 GATCAACACAGCTGACCTCTTTGTC----- 275
Db 475 AsparGAspLysGluGlnLeuCysAsnIleMetAspTyrGlnArgLeuSerIleAspAla 494
Qy 274 -----GCTTCA----- 251
Db 495 CysAlaHisAlaSerHisAsnAspArgLeuProLeuArgValValLeuGlnValLeuPhe 514
Qy 250 ATGCAAGCCCTCCACCTCCAGGGCTCTTGCTTAGGTGGCTGTAGCATCCCTACCAC 191
Db 515 PheGluGlnMet-HisLeuArgThrAlaLeuAla--GlyGlyLeuAsnValAlaAsnTh 533
Qy 190 CCAGGACACTGCTCGAATGACAACTCAAGTTGGAGGGGGAACAGGGAAGAGGAT 131
Db 533 rGIuThr-----AlaHisAlaValThrIleProGlyGlyArgThrGlyGlnGluIleVa 551
Qy 130 GGATGGGGTGGTGTATCTTACTCTGTTTAAAGCAGAACACCTTGTTGCGGTG 78
Db 551 iGlnArgAspGly-TrpValThrValValArgGlnAsnGlnValLeuLysVal 568
RESULT 10
JE0211
hypothetical 44k protein - Bombyx mori cytoplasmic polyhedrosis virus strain I
C;Species: Bombyx mori cytoplasmic polyhedrosis virus, BmCPV
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 20-Jun-2000
C;Accession: JE0211
R;Hagiwara, K.; Tomita, M.; Kobayashi, J.; Miyajima, S.; Yoshimura, T.
Biochem. Biophys. Res. Commun. 247, 549-553, 1998
A;Title: Nucleotide sequence of Bombyx mori cytoplasmic polyhedrosis virus segment 8.
A;Reference number: JE0211; MUID:98321164; PMID:9657006
A;Accession: JE0211
A;Molecule type: mRNA
A;Residues: 1-390 <HAG>
A;Cross-references: GB:AB016436; NID:g3395150; PIDN:BAA32042.1; PID:g3395151
Alignment Scores:
Pred. No.: 4.49 Length: 390
Score: 81.00 Matches: 27
Percent Similarity: 49.02% Conservative: 23
Best Local Similarity: 26.47% Mismatches: 44
Query Match: 7.33% Indels: 8
DB: 2 Gaps: 4
US-09-914-191-1 (1-598) x JE0211 (1-390)
Qy 577 GTCAAGCCCAAGGAGGATACAGGAACTCTTNCAGCAGACAGAATGAGATTTCGTT 518
Db 136 IleGluSerGluGlyGluValAsnAspThrProGlnAspAspGluAlaGluLysLysAsp 155
Qy 517 TCTGAATTCAACTCAACTAATAAT-----GAGCTGCCGATGTATAGGAAAGGG 467
Db 156 ValGluIleArgMetAsnTrpSerGluGluIleValGluLeuProArgGlnGluGlu 175
Qy 466 ACTGTGTGTATATGCGAGAGGTGGATGAGTGTATGATGACAAAGAAATAGTCCCAACA 407
Db 176 AsnValLeuValLeuSerLysProSerMetIleSerGluGluGluLeu---MetProThr 194
Qy 406 GAAATGAAGGAAAAAGATGGCAGTGCACCGGACAGGACAAAGCCAGTGCCTTCAC 347
Db 195 AspMetGlu-----ValAlaThrProLysValLeuGluProProThrProLeuPro 211
Qy 346 TGCATATCATCGGGATGCTTTCTGGAAGGAACATCCAGAGATTCCTAGATGAGACAGC 287
Db 212 SerProIleIle---ValAlaValSerSerGluSerProGlnValLysGluIleGluArg 230
Qy 286 TGACCC 281
Db 231 ProPro 232
RESULT 11
AF2172
membrane-bound lytic transglycosylase A alr2933 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AF2172
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata,
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2172
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-410 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA174632.1; PID:g17132027; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2933
Alignment Scores:
Pred. No.: 5.1 Length: 410
Score: 80.50 Matches: 44
Percent Similarity: 35.33% Conservative: 15
Best Local Similarity: 26.35% Mismatches: 44
Query Match: 7.29% Indels: 64
DB: 2 Gaps: 7
US-09-914-191-1 (1-598) x AF2172 (1-410)
Qy 478 TATAGGAAGGGACTGTGTGATATGCGAGAAGGTGGATGAGTGTACAAAGAAATT 419
Db 251 TyrAlaGlyGlyThrAspTyrProThrThrSerIleGlyArgGluLeuAlaLysAspGly 270
Qy 418 AAGCTGCCA---ACAGAAATGGAAGAAAAAGATGGCAGTGCACCCGGACAGGACAAAG 362
Db 271 LysLeuProLeuSerGlyMetThrMetProLysMetIle----- 283
Qy 361 CCACTGCCCTTGACATGCGATATCATCGGGGATGCTTCTGGGAAGAACATCCAGAGATT 302
Db 284 -----SerPhePheArgGlnGlnProGluGlu 292
Qy 301 CTA-----GATGAAGACAGACTGA 284
Db 293 LeuAsnAsnTyrLeuProArgTrpGluArgPheValPhePheGlnGlnThrGlySerArg 312
Qy 283 CCCTTTGCGCTTCAGTCTCTGGTGTCTTAACCATGCAA----- 245
Db 313 ProAlaThrGlySerIleAsnValProValThrProGluArgSerIleAlaThrAspLys 332
Qy 244 GCCTCCACCTCCAGGGCTCCTTCCTTAGTGGCTGTAGCATCCCT----- 196
Db 333 SerLeuMetPro-ProGlyAlaLeuAlaLeuValHisThrSerIleProPheProValAl 352
Qy 195 -----ACCACCCAGCACACTGGTGC 176
Db 352 aGlyGlyArgMetGluTyrArgThrValSerArgTyrValLeuAspGlnAspThrGly-- 371
Qy 175 GAATGACACAACTCAAGTTGGGAGG-----GGAAACAGGAAGGAAGGAT 131
Db 372 -AsnAlaIleLysGlyProGlyArgValAspTyrPheMetGlyThrGlyLysGlnAlaGl 391
Qy 130 GGATGGGGTGGTGTATCT 112
Db 391 yAspArgAlaGlyValThr 397
RESULT 12
T18319
hypothetical protein L7610.10 - Leishmania major
C;Species: Leishmania major
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C;Accession: T18319
R;Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.; Ivens, A.;
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18876
A;Accession: T18319

A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-320 <OLI>
 A;Cross-references: EMBL:AL034356; NID:e1371878; PIDN:CAA22245.1
 C;Genetics:
 A;Note: L7610.10

Alignment Scores:

Pred. No.: 7.17 Length: 320
 Score: 79.00 Matches: 44
 Percent Similarity: 31.98% Conservative: 11
 Best Local Similarity: 25.58% Mismatches: 52
 Query Match: 7.21% Indels: 65
 DB: 2 Gaps: 7

US-09-914-191-1 (1-598) x T18319 (1-320)

Qy 78 CACGCAACAAAGGTGTTCTGCTTAAACAGAGTAAGATACAC---CACCCCAT----- 128
 Db 69 HisArgLeuCySerLeuLeuSerLeuLeuLeuHisThrHisProHisIleThr 88
 Qy 128 ----- 128
 Db 89 TyrValHisThrSerThrArgAsnHisLeuSerLeuCysValPheLeuSerGlyLysAsp 108
 Qy 129 -----CCATCCCTTCTTCCCTGTTCCCTCC 155
 Db 109 AlaThrSerValArgAlaProSerProProSerProProSerLeuHisProSer 128
 Qy 156 CAACCTGAGTTGTGTCATTCCGACACAGTCTCTGGGTGGTAGGATGCTACGCCACCTA 215
 Db 129 ArgLeu-----CysTyr----- 132
 Qy 216 AGGCAAGAGCCCTGGAGGTGGAGGGCTTGC-----ATGTTAAGCACACAGAACT 269
 Db 133 -----GlyArgCysArgSerAlaAlaSerHisLeuAlaGlu 145
 Qy 270 GAAGCGCAAAAGGTGAGTCTTCATCTAGATCTCTGATCTTCTCCAGAAAGCA 329
 Db 146 AlaGlyArgGlyGlnIleAlaArgGluArgLysGlyThrGlyArgGAsp 165
 Qy 330 TCCCGCATGATATCGCAGTGCAGGGCACTGGCTTTCTCTGGTCC-----GGGTC 380
 Db 166 GlyProSerProProSerCysAlaCysPheSerPheValLeu-SerValAlaGluThrTh 185
 Qy 381 ACTGCCATCTTTTCTTCATCTCTGTTGGCAGCTTAATTTCTTTGTCATCACTCA 440
 Db 185 rLeuProProMetArgSerPro-----SerAlaProProLeuLeuSerProPr 203
 Qy 441 TCCACCTTCTGCCATATCAACACAGTCCCTTCC 474
 Db 203 oProProSerSerThrProProSerGlnLeuAla 214

RESULT 13

T31614

hypothetical protein Y50E8A.j - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T31614

R;Steward, C.

submitted to the EMBL Data Library, September 1999

A;Reference number: 221047

A;Accession: T31614

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-225 <WIL>

A;Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55053.1; CESP:Y50E8A.j

A;Experimental source: clone Y50E8A

C;Genetics:

A;Gene: CESP:Y50E8A.j

Alignment Scores:

Pred. No.: 7.8 Length: 225

Score: 78.50 Matches: 33
 Percent Similarity: 37.86% Conservative: 20
 Best Local Similarity: 23.57% Mismatches: 37
 Query Match: 7.16% Indels: 50
 DB: 2 Gaps: 7

US-09-914-191-1 (1-598) x T31614 (1-225)

Qy 220 AAGAGCCCTGGAGGTGGGA-----GGGCTTCATCG 252
 Db 109 LysHisProTrpAspProGlyThrGlnGlnArgTyrGlnLeuAlaProArgThrGlnTrp 128
 Qy 253 TTAAGCACACCAAGACTGAAGCGCAAAAGGTCAGCTGCTTCATCTAGAACTCTCGGAT 312
 Db 129 MetProThrProLeuLeuAspGlnAlaAlaIle-LeuSerPheLeuSerIleLeuTyrIle 148
 Qy 313 GTTCCTTCCAGAAAGCATCCCGATGATATCCAGTGCAGGGCACTGGCTTTGTCCTGG 372
 Db 148 ephe-----SerIleProHis-----G1 154
 Qy 373 TCGGGTCACTGCCATCTTTTCTTCCATTTCTGTCGACGCTT-----AA 420
 Db 154 yProSerThrCysAsnValPheSerIleTrpProGlnTrpAsnPheGlyAsnSerAsnAs 174
 Qy 421 TTTCCTT-----TGTCACTACCTTCCACCTTCTGCCATATCAACA 462
 Db 174 nTyrTrpMetProIleGlyGlnGlySerHisPheSerTyrLeuLeuProGlnSe 194
 Qy 463 CAGTCCCTTTCCTATACATCGGAGCTCATATTATAGTTGATGTGAATTGAGAAACA 522
 Db 194 rAla-----AlaAsnLeuGlnPheArgLeuPr 203
 Qy 523 AAATCTCATCTTGTCTGC-----TGNAAGAGTTCCTCTGAATCTCCC 565
 Db 203 oGluPheSerMetIleCysTyrGlnArgTyrGlnIleLysSerThrMetAspCysPro 222

RESULT 14

H84071

lipopolysaccharide biosynthesis BH3376 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: H84071

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Fuji, F.; Hi:

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; WUID:20512592; PMID:11058132

A;Accession: H84071

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-397 <STO>

A;Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BAB07095.1; GSPDB:GNO:

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH3376

Alignment Scores:

Pred. No.: 8.29 Length: 397
 Score: 78.50 Matches: 35
 Percent Similarity: 40.98% Conservative: 15
 Best Local Similarity: 28.69% Mismatches: 35
 Query Match: 7.16% Indels: 37
 DB: 2 Gaps: 8

US-09-914-191-1 (1-598) x H84071 (1-397)

Qy 19 TATAAATAAGTACTCGGATTAAAAAAGCACTT-----CTG 57

Db 149 TrpLysTyrSerThrGluPheGluSerArgAlaLeuSerGlnHisIleIleThr 168

Qy 58 CCAAGGAACCATGTTCCCAACACCGCCAAACAGGTTCTGCTTAACAGACAGTAACATAC 117

Db 169 ProSerAsnTrpLeuAsnHisTyrLeuLysGluLeuPheCysLeuLysGlyValArgIle 188

118	ACCACCCCATCCATCCCTTCCCTCCCTGTGCC-----	150
	:	
189	ThrThr---ValHisAsnGlyLeuSerValSerPheLeuSerLysLeuAsnGlnAla	207
151	-----CTCCCCAAC-----TTGAGTTGTGTCACTTCGCACCCAGT	183
	:	
208	PheHisProProSerGlnSerLysLysValLeuValLeuSerCysIleAlaArgLeuThr	227
	:	
184	GTCTCGGTGTAGG-----GATGCTACAGCCACCTAAGCAAGGAG-----	225
228	ProLeuLysGlyHisIleTyrLeuLeuAspAlaLeuAlaGluLeuLysLysAsnThrLeu	247
226	CCCTCGGAGGTGGGAGGCTTGCATGG---TTAAGCACACAGACATGAAGCGCAAAAGG	282
	:	
248	AspTyrGlu-----CysTyrPheIleGlyAsnGlyGluIleLysLysLysLeu	263
283	GTACAGC	288
264	ValAsn	265

RESULT 15
Tl8274
1-l-phosphatidylinositol 3-kinase (EC 2.7.1.137) 3 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl8274
R:Zhou, K.; Takegawa, K.; Emr, S.D.; Firtel, R.A.
Mol. Cell. Biol. 15, 5645-5656, 1995
A:Title: A phosphatidylinositol (PI) kinase gene family in Dictyostelium discoideum: Bld
A:Reference number: Z06411
A:Accession: Tl8274
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1585 <ZHO>
A:Cross-references: EMBL:U23478; NID:g733523; PID:g733524; PIDN:AAA85723.1
C:Genetics:
A:Gene: PIK3
C:Keywords: phosphotransferase

Alignment Scores:	
Pred. No.:	9.62
Score:	78.50
Percent Similarity:	63.99%
Best Local Similarity:	23.29%
Query Match:	7.10%
DB:	2
Gaps:	10
Indels:	76
Mismatches:	62
Conservative:	30
Matches:	51
Length:	1585

US-09-914-191-1 (1-598) x T18274 (1-1585)

571	QY	GCCCAAGGGAGGATTCACAGGGAACTCTTNCACGACACAGAATAATCAGATTTTGTTTCT---	515
		: : : : : : : : : : : : : : :	
498	D _B	AlaserGlyLysMetGlnGlyHisLeuAsnLeuGluLysAspGluLysPhePheThrLeu	517
		: : : : : : : : : : : : : : :	
514	QY	-----GNA 512	
		-----: : :	
518	D _B	ArgTrpCysAsnLysAspValValPheAspGlnAspThrProLeucGlyHisLeulleGln	537
		: : : : : : : : : : : : : : :	
511	QY	TTCACATCAACTATTAATGACTCCGCGATGTATAGGAAGGGACTGTGTGATATGG	452
		: : : : : : : : : : : : : : :	
538	D _B	TyrAsnLeuAsnTyrrAsnAsn-----ProThrGlnLysProThrAsnile	552
		: : : : : : : : : : : : : : :	
451	QY	CAGAAGGTGGATGAAGTGCATCACAAAAGAAATTAAAGCTGCCAACAGAAATGGAAGGAAAA	392
		: : : : : : : : : : : : : : :	
553	D _B	---lysleuGluLeuValLeuLysGluLeuCysLysGluArgLeuValAspLeuGln	571
		: : : : : : : : : : : : : : :	
391	QY	AAGATGGGACGATCCCAGCACAGACAAAAGCCAGTCCTTGCCACTCCGATATCATCCGG	332
		: : : : : : : : : : : : :	
572	D _B	SerLeuGluIleAsnAsnGlyArg-----	579
		: : : : : : : : : : : : : : :	
331	QY	GATGCTTCTCGAAGGAACAT---CCAGAGATTCATAGATGAAGACAGCTGACCCTTTTGC	275
		: : : : : : : : : : : : : : : : : :	
580	D _B	ProSerIleTrpLysSerHisIleAspAspValLeuSerPheAsn-----	594
		: : : : : : : : : : : : : : :	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 9, 2003, 12:45:33 ; Search time 13 Seconds
(without alignments)
3815.822 Million cell updates/sec

Title: US-09-914-191-1
Perfect score: 1096
Sequence: 1 ttggaatagttctgtctta.....ggtgttagccagattgttg 598

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_epool/US09914191/runat_09072003_112348_12866/app_query.fasta_1.775
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09914191 @CGN 1.1 26 @runat_09072003_112348_12866 -NCFU=6 -ICPU=3
-NO_MMAP -LARGESOURCY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	122	11.0	237	1 YG1G YEAST	P53215 saccharomyc
C 2	85.5	7.8	495	1 ENP2 RAT	Q35795 rattus norv
C 3	84.5	7.7	495	1 ENP2 MOUSE	O55026 mus musculu
C 4	80	7.2	304	1 PUR7 PICPA	Q3C114 pichia past
C 5	80	7.2	673	1 FPO3 HUMAN	Q43524 homo sapien
C 6	79.5	7.3	2144	1 CLR2 RAT	Q9GY22 rattus norv
C 7	78.5	7.1	1585	1 P3K3 D1CD1	P54675 dictyosteli
C 8	78	7.1	700	1 DTNB MOUSE	O70585 mus musculu
C 9	78	7.1	743	1 DTNA HUMAN	Q9Y418 homo sapien
C 10	78	7.1	862	1 SLA2 BACAN	P94217 bacillus an
C 11	78	7.1	2907	1 FEN2 MOUSE	Q61555 mus musculu
C 12	77	7.0	627	1 DTNB HUMAN	O60941 homo sapien
C 13	76	6.9	304	1 GAL PSEFL	P11886 pseudomonas
C 14	76	6.9	490	1 MYOC MOUSE	O70624 mus musculu
C 15	76	6.9	1570	1 P3K1 D1CD1	P54673 dictyosteli
C 16	75.5	6.9	512	1 FEN2 YEAST	P25621 saccharomyc
C 17	75.5	6.9	615	1 CPO DROME	Q01617 drosophila
C 18	75	6.8	415	1 YLN2 CAEEL	Q18964 caenorhabdi

C 19	74	6.7	469	1 FXGA HUMAN	P55316 homo sapien
C 20	73.5	6.7	544	1 YN66 TREPA	O83932 treponema p
C 21	73.5	6.7	591	1 MNT MOUSE	O08789 mus musculu
C 22	73	6.6	502	1 MYOC RAT	O91114 rattus norv
C 23	73	6.6	3010	1 POLG HCVJA	P26662 h genome po
C 24	72.5	6.6	551	1 AGLA RHIME	O92318 rhizobium m
C 25	72.5	6.6	1013	1 PTPX MACNE	O02695 macaca neme
C 26	72.5	6.6	3010	1 POLG HCVTW	P29846 h genome po
C 27	72	6.6	141	1 YPRO OWEFU	P21260 owenla fusi
C 28	72	6.6	684	1 SKIL HUMAN	P12757 homo sapien
C 29	71.5	6.5	395	1 HOCI YEAST	P47124 saccharomyc
C 30	71.5	6.5	611	1 ILVD AGRTS	Q8ue43 agrobacteri
C 31	71	6.4	247	1 APMI MOUSE	O60994 mus musculu
C 32	71	6.4	331	1 ODPB PORPU	P51266 porphyra pu
C 33	71	6.5	614	1 NRDI HUMAN	P20393 homo sapien
C 34	71	6.5	1473	1 OVOS CHICK	P20740 gallus gall
C 35	71	6.5	2911	1 FBN2 HUMAN	P35556 homo sapien
C 36	71	6.5	2923	1 CLR2 HUMAN	Q9hcu4 homo sapien
C 37	70.5	6.4	306	1 PUR7 YEAST	P27616 saccharomyc
C 38	70.5	6.4	478	1 CYCA GLUOX	Q47945 gluconobact
C 39	70.5	6.4	1124	1 POL FIVT2	P31822 feline immu
C 40	70.5	6.4	3010	1 POLG HCVBK	P26663 h genome po
C 41	70.5	6.4	3010	1 POLG HCVJT	Q00269 h genome po
C 42	70	6.3	291	1 PUR7 CANEA	P27602 candida mal
C 43	70	6.4	386	1 SHIB SPAEH	P56496 spalax leuc
C 44	70	6.4	478	1 OPN4 HUMAN	Q9uhm6 homo sapien
C 45	70	6.3	776	1 TFR1 CHICK	Q90997 gallus gall

ALIGNMENTS

RESULT 1

YG1G YEAST
ID YG1G YEAST STANDARD; PRT; 237 AA.
AC PS3215;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 27.8 kDa protein in VMA7-RPS25A intergenic region.
GN YGR024C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RX Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII.";
RL Yeast 13:1077-1090(1997)
CC -!- SIMILARITY: TO S.POMBE SPCC63.07.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z72809; CAA97007.1; --
DR SGD; S0003256; YGR024C.
DR Hypothetical protein.

SK SEQUENCE 237 AA; 27757 MW; CCB018E24B9BEA26 CRC64;

Alignment Scores:

Pred. No.: 0.000128 Length: 237
Score: 122.00 Matches: 23
Percent Similarity: 75.00% Conservative: 10
Best Local Similarity: 52.27% Mismatches: 11
Query Match: 11.04% Indels: 0

Db 363 ValGlyThrLeuLysGlnLeuGlu-----GluAlaThrGluLeuThrCysAsnGlnThr 380

Oy 567 TGG 569
|||
Db 381 Trp 381

RESULT 3

ENP2_MOUSE
ID ENP2_MOUSE STANDARD; PRT; 495 AA.
AC O55026; O35928; Q9DCR9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 2 (EC 3.6.1.3)
DE (NTPDase2) (ecto-ATPase) (CD39 antigen-like 1).
GN ENTPD2 OR CD39LI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. (SHORT FORM).
RP TISSUE=Embryo;
RC MEDLINE=97419269; PubMed=9271669;
RA Chadwick B.P., Frischaut A.-M.;
RT "Cloning and mapping of a human and mouse gene with homology to ecto-ATPase genes.";
RL Mamm. Genome 8:668-672(1997).
[2]
SEQUENCE FROM N.A. (LONG FORM).
RP TISSUE=Hepatoma;
RC MEDLINE=98288263; PubMed=9624117;
RA Gao L., Dong L., Whitlock J.P. Jr.;
RT "A novel response to dioxin. Induction of ecto-ATPase gene expression.";
RL J. Biol. Chem. 273:15358-15365(1998).
[3]
SEQUENCE FROM N.A. (LONG FORM).
RP STRAIN=C57BL/6J; TISSUE=Kidney;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baren G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: IN THE NERVOUS SYSTEM, COULD HYDROLYZE ATP AND OTHER
CC NUCLEOTIDES TO REGULATE PURINERGIC NEUROTRANSMISSION. HYDROLYZES
CC ADP ONLY TO A MARGINAL EXTENT (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate;
CC -!- COFACTOR: REQUIRES CALCIUM AND MAGNESIUM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- INDUCTION: BY DIOXIN.
CC -!- PTM: HAS PROBABLY DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO THE GDAL / CD39 NTPASE FAMILY.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U91511; AAB81014.1; -;
DR EMBL; AF042811; AAC24347.1; -;
DR EMBL; AK002553; BAB22182.1; -;
DR MGD; MGI:1096863; Entpd2
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39; 1.
DR PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
KW Hydrolase; Transmembrane; Glycoprotein; Calcium; Magnesium;
KW Alternative splicing.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 POTENTIAL.
FT DOMAIN 26 462 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 463 483 POTENTIAL.
FT DOMAIN 484 495 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 130 132 LTS -> MAG (IN SHORT ISOFORM).
FT CONFLICT 133 495 MISSING (IN SHORT ISOFORM).
FT CONFLICT 400 400 T -> A (IN REF. 3).
FT CONFLICT 414 414 S -> R (IN REF. 3).
FT CONFLICT 437 437 T -> A (IN REF. 3).
SQ SEQUENCE 495 AA; 54310 MW; CC0811C5D79CA4C8 CRC64;

Alignment Scores:

Pred. No.: 1.12 Length: 495
Score: 84.50 Matches: 42
Percent Similarity: 37.80% Conservative: 20
Best Local Similarity: 25.61% Mismatches: 69
Query Match: 7.71% Indels: 33
DB: 1 Gaps: 9

US-09-914-191-1 (1-598) x ENP2_MOUSE (1-495)

Oy 87 CAAGGTGTTCTGTTAAACAGAGTAGATACACACCCCATCCCTTCTCCCTG 146
Db 250 GlnArgLeuLeuAlaSerAlaLeuGlnIleHisArgPheHisProCysTrpProLysGly 269
Oy 147 TTCCCTCCCAACTTGAGTTGTGTATTCGACACAGTGTCTCTGGTGGTAGGATCTAC 206
Db 270 TyrSerThrGlnValLeuArgGluValTyrGlnSerPro-----CysThr 285
Oy 207 AGCCACTAAGGCAAGGACCCCTGGAGGTGGAGGCTTGATGTTAAGCACACCAGA 266
Db 286 MetGlyGlnArgProGlnThrPheAsnSerSerAlaThrValSerLeuSerGlyThrSer 305
Oy 267 ACTGAAGCCCAAAAGGTGAGTGTCTTCATCT-----AGAATCTCTGATGTTCTCTC 320
Db 306 AsnAlaAlaLeuCysArgAspLeuValSerGlyLeuPheAsnIleSerSerCysProPhe 325
Oy 321 CAGAAAGCATCCCGCATGATATCGCAGTGCAGAGGCATGGCTTTTGTCTCGTCCGGGTC 380
Db 326 -----SerGlnCysSerPheAsnGly----- 332
Oy 381 ACTGCCATCTTTTCTTCATTTCTGTGTGCAGCTTAATTTCTTTTGTCACTTCA 440
Db 333 -----ValPheGlnProProValAla---GlyAsnPheIleAlaPhe----- 345
Oy 441 TCCACCTCTGTCATATCAACAGTCCCTTTCTTATATACATCGGAGCTCATTATTATAG 500
Db 346 SerAlaPheTyrTyr-----ThrValAspPheLeu-----LysThrValMetGly 360
Oy 501 TTGATGTTGAATTCAGAAACAAATCTCATTTCTGTCTGCTGNAAGAGTTCCTCGTAAT 560

Db 361 LeuProValGlyThrLeuLysGlnLeuGlu-----AspAlaThrGluThrThrCysAsn 378
 QY 561 CTCCTTGGGCT 572
 Db 379 GlnThrTrpAla 382

RESULT 4

PUR7_PICPA STANDARD; PRT; 304 AA.
 AC Q9C1J4;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phosphoribosylamidimidazole-succinocarboxamide synthase (EC 6.3.2.6)
 DE (SAICAR synthetase).
 GN ADE1.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]

SEQUENCE FROM N.A.
 RX MEDLINE=21125227; PubMed=1123254;
 RA Lin Cereghino G.P., Lin Cereghino J., Sunga A.J., Johnson M.A.,
 RA Lim M., Gleeson M.A.G., Clegg J.M.;
 RT "New selectable marker/auxotrophic host strain combinations for
 RT molecular genetic manipulation of Pichia pastoris.";
 RL Gene 263:159-169(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + 5-amino-1-(5-phospho-D-
 CC ribosyl)imidazole-4-carboxylate + L-aspartate = ADP + phosphate +
 CC (S)-2-[5-amino-1-(5-phospho-D-ribosyl)imidazole-4-
 CC carboxamido]succinate.
 CC -1- PATHWAY: De novo purine biosynthesis; seventh step.
 CC -1- SIMILARITY: BELONGS TO THE SAICAR SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL; AF321096; AA06766.1; -
 DR HSP; P27616; IA48.
 DR InterPro; IPR001636; SAICAR_synt.
 DR Pfam; PF01259; SAICAR_synt; 1.
 DR ProDom; PD003043; SAICAR_synt; 1.
 DR TIGRFAMs; TIGR00081; purG; 1.
 DR PROSITE; PS01057; SAICAR_SYNTHETASE_1; 1.
 DR PROSITE; PS01058; SAICAR_SYNTHETASE_2; 1.
 KW Ligase; Purine biosynthesis.
 SQ SEQUENCE 304 AA; 34101 MW; 584073DA712262C9 CRC64;

Alignment Scores:

Pred. No.: 3.03 Length: 304
 Score: 80.00 Matches: 29
 Percent Similarity: 40.94% Conservative: 23
 Best Local Similarity: 22.83% Mismatches: 37
 Query Match: 7.24% Indels: 38
 DB: 1 Gaps: 6

US-09-914-191-1 (1-598) x PUR7_PICPA (1-304)

QY 589 GGACTACACAGTACAGCCAGGAGGATTACAGGAACTCTTCAGCAGACAGAAT 530
 Db 10 GlyileLeuProLeulleAlaLysGlyLysAlaValArgAspIleTrAlaValAspGluAsn 29
 QY 529 GAGATTTTGT-----TCTGAATTCACATCAACTATATAATAG 488
 Db 30 AsnLeuLeuPheValAlaThrAspArgIleSerAlaTrpAspValIleMetThrAsnGly 49

QY 487 CTGCCGATGTATAGGAAGGACTGTGTTGATATGCAGAGGTG----- 443
 Db 50 ilePro-----AspLysGlyLysIleLeuThrGlnLeuSerValPheAspPhe 67
 QY 442 -----GATGAAGTGTATGACA--- 428
 Db 68 LeuAlaProTyrlleLysAsnHisLeuValAlaSerAsnAspLysGluValPheAlaLeu 87
 QY 427 -----AAAGAAATTAACTGCCACAGAAATGGAAGGAAAAAGATGGCA 383
 Db 88 LeuProSerLysLeuSerGluLysTyrlLysSerGlnLeuGluGlyArgSerLeulle 107
 QY 382 GTGACCCGACAGCAGCAAGCCAGTGCCTTGCACTGCATATCATCGG----- 332
 Db 108 ValLysLysHisArgLeu-----IleProLeuGluAlaIleValArgGlyTyrlleThr 125
 QY 331 GATGCTTTCTGGAAGGACAT 311
 Db 126 GlySerAlaTrpLysGluTrp 132

RESULT 5

FX03_HUMAN STANDARD; PRT; 673 AA.
 ID FX03_HUMAN
 AC O43524; O15171;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Forkhead box protein O3A (Forkhead in rhabdomyosarcoma-like 1) (AF6q21
 DE protein).
 GN FOXO3A OR FKHL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=98140118; PubMed=9479491;
 RA Anderson M.J., Viars C.S., Czekay S., Cavenee W.K., Arden K.C.;
 RT "Cloning and characterization of three human forkhead genes that
 RT comprise an FKHR-like gene subfamily.";
 RL Genomics 47:187-199(1998).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RX SEQUENCE OF 1-383 FROM N.A.
 RX MEDLINE=98008138; PubMed=9345057;
 RA Hillion J., Le Coniat M., Jonveaux P., Berger R., Bernard O.A.;
 RT "AF6q21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines
 RT a forkhead transcriptional factor subfamily.";
 RL Blood 90:3714-3719(1997).

CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT MAY TRIGGER
 CC APOPTOSIS BY INDUCING THE EXPRESSION OF GENES THAT ARE CRITICAL
 CC FOR CELL DEATH.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: UBIQUITOUS.

CC -1- PTM: PHOSPHORYLATED BY AKT1.

CC -1- DISEASE: INVOLVED IN A T(6;11)(Q21;Q23) CHROMOSOMAL TRANSLOCATION
 CC IN SECONDARY ACUTE LEUKEMIA THAT INVOLVES FOXO3A AND MLL.

CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;

CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF6q21D125.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----


```
DR EMBL; AF032886; AAC39592.1; --
DR EMBL; BC020227; AAH20227.1; --
DR EMBL; BC021224; AAH21224.1; --
DR EMBL; AJ001589; CAA04860.1; --
DR EMBL; AJ001590; CAA04861.1; --
DR HSSP; O63245; 2HFH.
DR TRANSFAC; T02938; --
DR Genew; HGNC:3821; FOXO3A.
DR MIN; 602681; --
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head.1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head.1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS0039; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Apoptosis;
KW Chromosomal translocation; Proto-oncogene; Phosphorylation.
FT DNA BIND 157 251 FORK-HEAD.
FT CONFLICT 156 163 AWGNLSYA -> WGNLPSYS (IN REF. 3).
FT CONFLICT 238 246 PDGKSGKA -> LMGERKT (IN REF. 3).
FT CONFLICT 253 253 S -> T (IN REF. 3).
FT CONFLICT 271 271 MISSING (IN REF. 3).
FT CONFLICT 292 330 PGSPTRSSDLDLAWTFDRGRTNSASTVSGRLSPIMAS
-> AWQPHVNAVMSWMRGRTSVHAPILTPAQSVAAACRPW
QV (IN REF. 3).
FT CONFLICT 345 361 PMLYSSSASLSPSVSKP -> AHALQHSQPVTFSKQA
(IN REF. 3).
FT CONFLICT 367 367 P -> R (IN REF. 3).
FT CONFLICT 371 371 D -> E (IN REF. 3).
FT CONFLICT 382 383 LT -> AD (IN REF. 3).
SQ SEQUENCE 673 AA; 71276 MW; ESB4E830665A9982 CRC64;

Alignment Scores:
Pred. No.: 3.43 Length: 673
Score: 80.00 Matches: 27
Percent Similarity: 40.40% Conservative: 13
Best Local Similarity: 27.27% Mismatches: 26
Query Match: 7.30% Indels: 33
DB: 1 Gaps: 2

US-09-914-191-1 (1-598) x FOXO3_HUMAN (1-673)
QY 87 CAAGGTGTTCTGCTTAACAGAGTAAGATACACACCCCATCCCTTCCTTCCTG 146
Db 525 GlnGlySerLeuValAsnGlnAsnLeuHisGlnHisGlnThr----- 540
QY 147 TTCCCTCCCAACTTGAGTTGTGTGTCATTCGCACACAGTGTCTCTGGGTAGGATGCTAC 206
Db 540 ----- 540
QY 207 AGCCACCTAAGGAGAGCCCTGGGAGGTGGGAGGCTTCATGTTAAGCACACAGAGA 266
Db 541 -----GlnGlyAlaLeuGlyGlySerArgAlaLeu-----SerAsn 552
QY 267 ACTGAAGGCAAAAGAGTGTCTCTCATCTAGAAATCTGTGATGTTCTCTCCAGAAA 336
Db 553 SerValSerAsnMetGlyLeuSerGluSerSerLeuGlySerAlaLysHisGlnGln 572
QY 327 GCATCCCGGATGATATCCAGTGCAGAGGCACTGGCTTGTCTCTGGTCCGGGTCA 381
Db 573 GlnSerProVal-SerGlnSerMetGlnThrLeuSerAspSerLeuSerGlySer 590

RESULT 6
CLR2_RAT ID CLR2_RAT STANDARD; PRT; 2144 AA.
AC Q9QYF2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal
DE growth factor-like domains 3) (Fragment).

GN CELSR2 OR MEGF3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.;
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. High expression in
CC cerebellum and olfactory bulb. Weaker expression in cerebral
CC cortex, hippocampus and brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: CONTAINS 4 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011529; BAA8687.1; --
DR HSSP; P00740; LEDM.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000832; GPCR secretin.
DR InterPro; IPR001879; horxn receptor.
DR InterPro; IPR002049; Laminin EGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR000203; PKC_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 3.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 3.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; HormR; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
DR PROSITE; PS00221; GPS; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
```

DR PROSITE; PS0261; G PROTEIN RECP F2 4; 1.
 KW PROSITE; PS01248; LAMININ TYPE_EGF_1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
 KW Developmental protein; Hydroxylation.
 FT NON TER 1 1
 FT DOMAIN 1 1605 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1606 1626 1 (POTENTIAL).
 FT DOMAIN 1627 1641 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1642 1662 2 (POTENTIAL).
 FT DOMAIN 1663 1663 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1664 1684 3 (POTENTIAL).
 FT DOMAIN 1685 1705 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1706 1726 4 (POTENTIAL).
 FT DOMAIN 1727 1744 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1745 1765 5 (POTENTIAL).
 FT DOMAIN 1766 1789 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1790 1810 6 (POTENTIAL).
 FT DOMAIN 1811 1816 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1817 1837 7 (POTENTIAL).
 FT DOMAIN 1838 2144 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN <1 40 CADHERIN 1.
 FT DOMAIN 41 146 CADHERIN 2.
 FT DOMAIN 147 248 CADHERIN 3.
 FT DOMAIN 253 371 CADHERIN 4.
 FT DOMAIN 453 511 EGF-LIKE 1, CALCIUM-BINDING.
 FT DOMAIN 513 549 EGF-LIKE 2, CALCIUM-BINDING.
 FT DOMAIN 553 591 EGF-LIKE 3, CALCIUM-BINDING.
 FT DOMAIN 592 796 LAMININ G-LIKE 1.
 FT DOMAIN 799 835 EGF-LIKE 4, CALCIUM-BINDING.
 FT DOMAIN 839 1016 LAMININ G-LIKE 2.
 FT DOMAIN 1018 1053 EGF-LIKE 5, CALCIUM-BINDING.
 FT DOMAIN 1054 1092 EGF-LIKE 6, CALCIUM-BINDING.
 FT DOMAIN 1108 1147 EGF-LIKE 7, CALCIUM-BINDING.
 FT DOMAIN 1148 1180 EGF-LIKE 8, CALCIUM-BINDING.
 FT DOMAIN 1541 1593 GPS.
 FT DOMAIN 1968 1973 POLY-GLU.
 FT DISULFID 457 468 BY SIMILARITY.
 FT DISULFID 462 499 BY SIMILARITY.
 FT DISULFID 501 510 BY SIMILARITY.
 FT DISULFID 512 528 BY SIMILARITY.
 FT DISULFID 522 537 BY SIMILARITY.
 FT DISULFID 539 548 BY SIMILARITY.
 FT DISULFID 557 568 BY SIMILARITY.
 FT DISULFID 562 578 BY SIMILARITY.
 FT DISULFID 580 590 BY SIMILARITY.
 FT DISULFID 803 814 BY SIMILARITY.
 FT DISULFID 808 823 BY SIMILARITY.
 FT DISULFID 825 834 BY SIMILARITY.
 FT DISULFID 1022 1033 BY SIMILARITY.
 FT DISULFID 1027 1042 BY SIMILARITY.
 FT DISULFID 1044 1053 BY SIMILARITY.
 FT DISULFID 1057 1068 BY SIMILARITY.
 FT DISULFID 1062 1080 BY SIMILARITY.
 FT DISULFID 1082 1091 BY SIMILARITY.
 FT DISULFID 1112 1124 BY SIMILARITY.
 FT DISULFID 1114 1131 BY SIMILARITY.
 FT DISULFID 1133 1146 BY SIMILARITY.
 FT DISULFID 1149 1161 BY SIMILARITY.
 FT DISULFID 1151 1168 BY SIMILARITY.
 FT DISULFID 1170 1179 BY SIMILARITY.
 FT MOD_RES 816 816 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 1035 1035 HYDROXYLATION (POTENTIAL).
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 966 966 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1249 1249 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1286 1286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1548 1548 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1570 1570 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2144 AA; 233480 MW; 6EA989C1BA655ECA CRC64;
 Alignment Scores:
 Pred. No.: 4.63 Length: 2144
 Score: 79.50 Matches: 43
 Percent Similarity: 34.02% Conservatives: 23
 Best Local Similarity: 22.16% Mismatches: 57
 Query Match: 7.25% Indels: 72
 DB: 1 Gaps: 12
 US-09-914-191-1 (1-598) x CLR2_RAT (1-2144)
 QY 16 CTTTATAAAATAGT-----ACTGCGATTAAAAAAGACATCTTGCCA----- 60
 Db 434 LeuTyrLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValLeuProPheAsp 453
 QY 61 -----AAGGAACCATGTTCCAAACCCGCAACAGGTGTCTGCTTAAC 105
 Db 454 AspAsnIleCysLeuArgGluProCysGluAsnTyrMetArg-----CysValSer 470
 QY 106 AGAGTAAGATACACACC-----CTTCCTCTGTTCCCTCC-----CCC 126
 Db 471 ValLeuArgPheAspSerSerAlaProPheIleAlaSerSerValLeuPheArgPro 490
 QY 127 ATCCATCCCTTC-----CTTCCTCTGTTCCCTCC----- 155
 Db 491 IleHisProValClyGlyLeuArgCysArgCysProProGlyPheThrGlyAspTyrCys 510
 QY 156 -----CAACTGAGTTGTCTATCGCACCATGTCCTGGGTGGTAGGATGCTACAGC 209
 Db 511 GluThrGluValAspLeuCysTyrSerArgProCys-----GlyProHisGly 526
 QY 210 CACCTAAGGCAAGGAGCCCTGGGAGGTGGAGGCTTGCATGGTTAAG----- 257
 Db 527 HisCysArg-----SerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 544
 QY 258 -----CACACCAGAACTGAAGCGCAAAAGGTGACGTCTTCTATCTAGATCTCTGGA 311
 Db 545 GlyGluHisCysGluValSerAlaArgSerGly----- 555
 QY 312 TGTTCCTCCAGAAACATCCCGATGATATCCAGTCCAGAGGCACTGCTTGTCTGCTG 371
 Db 556 -----ArgCysThrProGlyVal-----CysLysAsnGlyThrCysVal 569
 QY 372 GTCCGGGTCACTGCCATCTTTTCTCTCCATTTCTGTGGCAGCTTAATTTCTTTTGTG 431
 Db 570 AsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPhe----- 585
 QY 432 ATCACTTCATCCACCTTTCGCCATATCAACACAGTCCCTTTC 473
 Db 586 -----GluLysProPheCysGlnValThrArgSerPhe 597
 RESULT 7
 P3K3_DICDI STANDARD; PRT; 1585 AA.
 ID P3K3_DICDI
 AC P54675;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-kinase)
 GN (Ptdins-3-kinase) (PI3K) (Fragment).
 GE PI3K OR PI3K.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX3;
 RX MEDLINE=96009592; PubMed=7565716;

RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
 RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
 RT discoidium: biological roles of putative mammalian p110 and yeast
 RL Mol. Cell. Biol. 15:5645-5656(1995)
 CC -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-ID-myo-inositol = ADP +
 CC 1-phosphatidyl-ID-myo-inositol 3-phosphate.
 CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U23478; AAA85723.1; -
 DR Dictyob; DD01101; pikC.
 DR InterPro; IPR002420; PI3K_C2.
 DR InterPro; IPR000341; PI3K_ras_bind.
 DR InterPro; IPR001263; PI3Ka.
 DR InterPro; IPR000403; PI3_P14_kinase.
 DR Pfam; PF00454; PI3_P14_kinase; 1.
 DR Pfam; PF00613; PI3Ka; 1.
 DR Pfam; PF00792; PI3K_C2; 1.
 DR Pfam; PF00794; PI3K_rbd; 1.
 DR SMART; SM00142; PI3K_C2; 1.
 DR SMART; SM00144; PI3K_rbd; 1.
 DR SMART; SM00145; PI3Ka; 1.
 DR SMART; SM00146; PI3K; 1.
 DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
 DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
 DR PROSITE; PS0290; PI3_4_KINASE_3; 1.
 DR TRANSFERASE; Kinase; Multigene family; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 58 84 POLY-ASN.
 FT DOMAIN 98 110 POLY-ASN.
 FT DOMAIN 200 226 POLY-ASN.
 FT DOMAIN 239 254 POLY-ASN.
 FT DOMAIN 345 378 POLY-ASN.
 FT DOMAIN 383 390 POLY-ASN.
 FT DOMAIN 484 488 POLY-SER.
 FT DOMAIN 720 737 POLY-GLN.
 FT DOMAIN 1569 1575 POLY-ASN.
 FT DOMAIN 1221 1484
 FT DOMAIN 1510 1539
 FT REPEAT 1510 1514
 FT REPEAT 1515 1519
 FT REPEAT 1520 1524
 FT REPEAT 1530 1534
 FT REPEAT 1535 1539
 FT REPEAT 1547 1560
 FT DOMAIN 1585 1595 7 X 2 AA TANDEM REPEATS OF K-E.
 SQ SEQUENCE 1585 AA; 180421 MW; 4689B620D2484961 CRC64;
 Alignment Scores:
 Pred. No.: 5.61 Length: 1585
 Score: 78.50 Matches: 51
 Percent Similarity: 36.99% Conservative: 30
 Best Local Similarity: 23.29% Mismatches: 62
 Query Match: 7.10% Indels: 76
 DB: 1 Gaps: 10
 US-09-914-191-1 (1-598) x PI3K_DICD1 (1-1585)
 QY 571 GCCCAAGGAGATACAGGAACTCTTNCAGCAGACAGAAATGAGATTTTGTCTTCT--- 515
 Db 498 AlaserglylsMetGlnGlyHisLeuAsnLeuGlulysAspGluLysPhePheThrLeu 517
 QY 514 -----GAA 512
 Db 518 ArgTrpCysAsnLysAspValValPheAspGlnAspThrProLeuGlyHisLeuLeuGln 537

QY 511 TTCAACATCACTATAATAATAGAGTCCGATGATATAGAAAGGAGGACTGTGTGATATGG 452
 Db 538 TyrAsnLeuAsnTyrAsnAsn-----ProThrGlnLysProThrAsnLeu----- 552
 QY 451 CAGAAGGTGGATGAAGTATGACAAAGAAATTAAGCTGCCCAACACAAATGGAGGAAAA 392
 Db 553 ---LysLeuGluLeuValLeuGluAspGluLeuCysLysGluArgLeuValAspLeuGln 571
 QY 391 AAGATGCGAGTACCCGGACAGGACAAAGCCAGTGCCTTGTGCACTGCGATATATCGGG 332
 Db 572 SerLeuGluLeuAsnAsnGlyArg----- 579
 QY 331 GATGCTTTCTGGAAGAAACAT---CCAGAGATTTCTAGTGAAGACAGCTGACCTTTTTC 275
 Db 580 ProSerIleTrpLysSerHisLeuAspValLeuSerPheAsn----- 594
 QY 274 GCTTTCAGTTCTGTGTGCTTAACCATGCAAGCCCTCCACCTCCAGGGCTCCTTGCCTT 215
 Db 595 ---ArgLysLeuArgGluLeuAlaMetLeuAlaLysProGlnSerAsnValProAlaAla 613
 QY 214 AGGTGGCTGTAGCATCCCTACACCCAGGACACTGGTGGCAATGACACAACTCAAGTTGG 155
 Db 614 ArgLeuThrProTyrPro--ProProLysThrIleProGluPhePhe----- 628
 QY 154 GAGGGGAACAGGGAAGGATGGATGGGGTGTGTATCTTCTTCTTCTTCTTCTTCTTCT 98
 Db 629 -----ValIleArgValHisLeuPheLysAsnG 638
 QY 97 AGAACACCTTGTGTGGGTGTTGGAC-----ATGGTTCCTTTG 59
 Db 638 In-ThrLysSerLeuArgCysAlaAsnAsnHisThrAlaPheSerLeuMetThrLeuLeu 657
 QY 58 GCAGAAAGTGTCTT-----TTTTTAAATCGCAGTACTATTTT 23
 Db 658 SerGluLysLeuLysAsnThrThrProPheAspProThrGlnTyrArgPhe 674
 RESULT 8
 DTNB_MOUSE STANDARD; PRT; 700 AA.
 AC 070585; 070563; Q9CTZ1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dystrobrevin beta (Beta-dystrobrevin) (DTN-B)
 GN DTNB.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=98200066; PubMed=9540997;
 RA Puca A.A., Piluso V.N.G., Belsito A., Sampaolo S., Quaderi N.,
 RA Rossi E., Di Iorio G., Ballabio A., Franco B.;
 RT "Identification and characterization of a novel member of the
 RT dystrobrevin gene family.";
 RL FEBS Lett. 425:7-13(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=98081858; PubMed=9419360;
 RA Blake D.J., Nawrothski R., Loh N.Y., Gorecki D.C., Davies K.E.;
 RT "Beta-dystrobrevin, a member of the dystrophin-related protein
 RT family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:241-246(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=99018217; PubMed=9799833;
 RA Loh N.Y., Ambrose H.J., Guay-Woodford L.M., Dasgupta S.,
 RA Nawrothski R.A., Blake D.J., Davies K.E.;
 RT "Genomic organization and refined mapping of the mouse beta-
 RT dystrobrevin gene.";
 RL Mamm. Genome 9:857-862(1998).
 RN [4]

RP SEQUENCE OF 590-608 FROM N.A. (ISOFORM 2).
 RC STRAIN=C57BL/6J; TISSUE=Stomach;
 RX MEDLINE=21085660; PubMed1217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 CC -1- SUBUNIT: INTERACTS WITH DYSTROPHIN SHORT FORM DP71 AND SYNTROPHIN.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN, KIDNEY, LIVER AND
 CC LUNG. IN BRAIN EXPRESSED IN NEURONS OF THE CORTEX AND HIPPOCAMPUS.
 CC -1- DOMAIN: THE COILED-COIL DOMAIN MAY MEDIATE THE INTERACTION WITH
 CC DYSTROPHIN.
 CC -1- SIMILARITY: BELONGS TO THE DYSTROPHIN FAMILY. DYSTROBREVIN
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y15742; CAA75752.1;
 DR EMBL; AJ003007; CAA05796.1;
 DR EMBL; AJ010204; CAA09038.1;
 DR EMBL; AJ010205; CAA09038.1; JOINED.
 DR EMBL; AJ010206; CAA09038.1; JOINED.
 DR EMBL; AJ010207; CAA09038.1; JOINED.
 DR EMBL; AJ010208; CAA09038.1; JOINED.
 DR EMBL; AJ010209; CAA09038.1; JOINED.
 DR EMBL; AJ010210; CAA09038.1; JOINED.
 DR EMBL; AJ010211; CAA09038.1; JOINED.
 DR EMBL; AJ010212; CAA09038.1; JOINED.
 DR EMBL; AJ010213; CAA09038.1; JOINED.
 DR EMBL; AJ010214; CAA09038.1; JOINED.
 DR EMBL; AJ010215; CAA09038.1; JOINED.
 DR EMBL; AJ010216; CAA09038.1; JOINED.
 DR EMBL; AJ010217; CAA09038.1; JOINED.
 DR EMBL; AJ010218; CAA09038.1; JOINED.
 DR EMBL; AJ010219; CAA09038.1; JOINED.
 DR EMBL; AJ010220; CAA09038.1; JOINED.
 DR EMBL; AJ010221; CAA09038.1; JOINED.
 DR EMBL; AK019668; BAB31331.1;
 DR MGD; MGI:1203728; Dtnb.
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00569; ZZ; 1.
 DR SMART; SM00291; Znf_ZZ; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS01357; ZF_ZZ_2; 1.
 KW Coiled coil; Zinc-finger; Alternative splicing.
 FT ZN FING 237 284
 FT DOMAIN 369 418
 FT SYNTROPHIN-BINDING REGION.

FT DOMAIN 428 521 COILED COIL (POTENTIAL).
 FT VARSPLIC 518 518 K -> KEEBOKA (IN ISOFORM 2).
 FT VARSPLIC 603 608 AEAEQ -> EVTPVS (IN ISOFORM 2).
 FT VARSPLIC 609 700 MISSING (IN ISOFORM 2).
 FT CONFLICT 412 412 P -> L (IN REF. 2 AND 3).
 FT CONFLICT 465 465 S -> F (IN REF. 2 AND 3).
 SQ SEQUENCE 700 AA; 78355 MW; 6BE34EB861AFB46C CRC64;
 Alignment Scores:
 Pred. No.: 5.57 Length: 700
 Score: 78.00 Matches: 42
 Percent Similarity: 35.59% Conservative: 21
 Best Local Similarity: 23.73% Mismatches: 53
 Query Match: 7.06% Indels: 61
 DB: 1 Gaps: 8
 US-09-914-191-1 (1-598) x DTNB_MOUSE (1-700)
 QY 538 GACAAGATGAGATTGTTTCTGAAATTCACATCAACTATAATAATGAGTCGCCGATG 479
 DB 146 ASPLYSLeuArgTyrIlePheSerGln-----Met 155
 QY 478 TATAGAAAGGAGCTGTGTATATGCGAAGGTGGATGAAGTGAATGACAAAGAAAT 419
 DB 156 SerAspSerAsnGlyLeuMetMetPheGlyLysLeuAspGlnPheLeuLysGluAlaLeu 175
 QY 418 AGCTGCCCAACAAATG---CAAGGAAAAAGATGGCAGTGACCCGGACCCAGGACAAAG 362
 DB 176 LysLeuProThrAlaValPheGluGlyProSerPheGlyTyrThrGluHisAlaValArg 195
 QY 361 CCA---GTCCCTTCGACATCATCGGGGATCTTCTCTGGAAGGAACATCCAGAG 305
 DB 196 ThrCysPheProGlnGlnLysLysIleMetLeuAsnMetPhe----- 209
 QY 304 ATTCTAGATGAAGACAGCTGACCCCTTTTGCCTTCAGTTCGTGTGTAAACATGCAA 245
 DB 210 ---LeuAsp-----ThrMetMet 214
 QY 244 GCCTCCCACTCCAGGGCTCTTCCCTTAGTGGCTG----- 206
 DB 215 AlaAspProProGln-----CysLeuValTyrLeuProLeuMethHisArgLeuAla 232
 QY 205 ---TAGCATCCCTACCACCCAGGACACTGGTGCATGACACAACTCAAGTTGGGAG 152
 DB 233 HisValGluAsnValPheHisProValGluCysSerTyrCysHisCysGluSerMetMet 252
 QY 151 GGAACAGGGAAGGAAGGATCGATGGGGGTGGTGTATCTTACTCTGTGTTAAGACAGAA 92
 DB 253 Gly-----PheArgTyr 256
 QY 91 CCTGTGTCGGGTGTGGNACATGGTTCCTTTGGCAGAGTCTTTTT 43
 DB 257 ArgCysGlnGlnCysHisAsn-TyrGlnLeuCysGlnAsnCysPheTyr 272
 RESULT 9
 DTNA HUMAN
 ID DTNA HUMAN STANDARD; PRT: 743 AA.
 AC Q9Y4J8; Q13498; Q13499; Q13199; Q13500; O15332; O75697;
 AC Q13197; Q13198;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dystrobrevin alpha (Dystrobrevin-alpha).
 GN DTNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 4; 5; 6; 7 AND 8).
 RX MEDLINE=96254978; PubMed=8845841;
 RA Sadoulet-Puccio H.M., Khurana T.S., Cohen J.B., Kunkel L.M.;
 RT "Cloning and characterization of the human homologue of a dystrophin


```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95263670; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.;
RT "Developmental expression of fibrillin genes suggests heterogeneity
of extracellular microfibrils.";
RL J. Cell Biol. 129:1165-1176(1995).
RN [2]
RP SEQUENCE OF 210-317 FROM N.A.
RX MEDLINE=9410368; PubMed=8307578;
RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J.,
FR Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on
mouse chromosomes 2 and 18.";
RL Genomics 18:667-672(1993).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39790; AAA74908.1; -.
DR EMBL; G69359; AAC60685.1; -.
DR HSPSP; P35555; 1EMN.
DR MGD; MGI:95490; Fbn2.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF; 46.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PR00010; EGFBL00D.
DR SMART; SM00179; EGF_CA; 43.
DR SMART; SM00001; EGF-like; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 43.
DR PROSITE; PS00022; EGF-1; 2.
DR PROSITE; PS01185; EGF-2; 36.
DR PROSITE; PS01187; EGF_CA; 43.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2907 FIBRILLIN 2.
FT DOMAIN 111 142 EGF-LIKE 1.
FT DOMAIN 145 176 EGF-LIKE 2.
FT DOMAIN 176 208 EGF-LIKE 3.
FT DOMAIN 276 317 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 318 359 EGF-LIKE 5, CALCIUM-BINDING.
FT REPEAT 360 426 TGFBP 1.
FT DOMAIN 487 527 EGF-LIKE 6.
FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 692 760 TGFBP 2.
FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 948 989 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 990 1065 TGFBP 3.
FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1566 1642 TGFBP 4.
FT DOMAIN 1643 1684 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1685 1726 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1727 1800 TGFBP 5.
FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2091 2163 TGFBP 6.
FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2373 2441 TGFBP 7.
FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2525 2563 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2564 2607 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 119 130 BY SIMILARITY.
FT DISULFID 132 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 196 BY SIMILARITY.
FT DISULFID 198 207 BY SIMILARITY.
FT DISULFID 280 292 BY SIMILARITY.
FT DISULFID 287 301 BY SIMILARITY.
FT DISULFID 303 316 BY SIMILARITY.
FT DISULFID 322 334 BY SIMILARITY.
FT DISULFID 329 343 BY SIMILARITY.
FT DISULFID 345 358 BY SIMILARITY.
FT DISULFID 491 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 526 BY SIMILARITY.
FT DISULFID 532 542 BY SIMILARITY.
FT DISULFID 537 551 BY SIMILARITY.
FT DISULFID 553 566 BY SIMILARITY.
FT DISULFID 572 584 BY SIMILARITY.
FT DISULFID 579 593 BY SIMILARITY.
FT DISULFID 595 608 BY SIMILARITY.
FT DISULFID 614 635 BY SIMILARITY.
FT DISULFID 620 634 BY SIMILARITY.
FT DISULFID 636 649 BY SIMILARITY.
FT DISULFID 655 666 BY SIMILARITY.
FT DISULFID 661 675 BY SIMILARITY.
FT DISULFID 677 690 BY SIMILARITY.
FT DISULFID 765 777 BY SIMILARITY.
FT DISULFID 772 786 BY SIMILARITY.
FT DISULFID 788 801 BY SIMILARITY.
FT DISULFID 807 819 BY SIMILARITY.
FT DISULFID 814 828 BY SIMILARITY.
FT DISULFID 830 843 BY SIMILARITY.
FT DISULFID 849 859 BY SIMILARITY.
FT DISULFID 854 868 BY SIMILARITY.
FT DISULFID 870 883 BY SIMILARITY.
FT DISULFID 952 964 BY SIMILARITY.

```



```
FT VARSPLIC 360 390 RLOYSQDIPSHLADEHALIASYVARLOHCAR -> S (IN
FT VARSPLIC 526 558 AQATGSPHTSTHGGGRMPMPVRSSTAGSTPT -> VSEL
FT VARSPLIC 559 608 RNFPPDSRSFARSSFPFHINVSPLPAHHL (IN ISOFORM
FT VARSPLIC 579 608 MISSING (IN ISOFORM 3).
FT VARSPLIC 608 MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ SEQUENCE 627 AA; 71355 MW; 6927EC0BD86D375C CRC64;

Alignment Scores:
Pred. No.: 6.95 Length: 627
Score: 77.00 Matches: 42
Percent Similarity: 31.41% Conservative: 18
Best Local Similarity: 21.99% Mismatches: 42
Query Match: 6.97% Indels: 89
DB: 1 Gaps: 9

US-09-914-191-1 (1-598) x DTNB_HUMAN (1-627)
QY 538 GACAGAATGAGATTGTTGTTTGAATTCACATCAACTATAATAATGAGCTCCGATG 479
DB 146 AsplysLeuArgTyrValPheSerGln-----Met 155
QY 478 TATAGGAAGGACTGCTTATATGCGAGAGGTGGATGATGATGACAAAAGAAATT 419
DB 156 SerAspSerAsnGlyLeuMetilePheSerLysPheAspGlnPheLeuLysGluValLeu 175
QY 418 AAGCTGCCAACAGAAATG---GAAGGAGAAAAAGATGGCAGTGCACCCGACCCAGGACAAAG 362
DB 176 LysLeuProThrAlaValPheGluGlyProSerPheGlyTyrThr-----190
QY 361 CCAGTGCCTTGCACCTGGATATCATCGGGGATGCTTCTGGAAGGAACATCCAGAGATT 302
DB 191 -----GluHis-----192
QY 301 CTAGATGAAGACACTGACCTTTGGCTTCAGTTCGTGGTGCTTA-----254
DB 193 -----SerValArgThrCysPheProGlnArg 202
QY 253 -----ACCATGAAGCCCTCCACCTCCAGCGGCTCT 221
DB 203 LysIleMetLeuAsnMetPheLeuAspThrMetMetAlaAspPropProGln-----220
QY 220 TGCCTTAGTGGCTG-----TAGCATCCTTACCACCCA 188
DB 221 CysLeuValTrpLeuProLeuMetHisArgLeuAlaHisValGluAsnValPheHisPro 240
QY 187 GGACAC-----TGGTGGGAATGACACAACTCAAGTTGGGAGGGAACAGGGAAGG 134
DB 241 ValGluCysSerTyrCysArg-----247
QY 133 GATGGATGGGGGTGTGTATCTTACTCTGTTTAAGCAGAACACCTTGTTCGGGTGTGG 74
DB 248 -----CysGluSerMetMetGlyPheArgTyrArgCysGlnGlnCysHis 262
QY 73 AACATGGTTCCTTTGGCAGAGTGCCTTTT 43
DB 263 Asn-TyrGlnLeuCysGlnAsnCysPheTrp 272

RESULT 13
GAL_PSEFL STANDARD; PRT; 304 AA.
AC P11886;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE D-galactose 1-dehydrogenase (EC 1.1.1.48).
GN GAL.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_taxID=294;
RN (1)
```

```
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC STRAIN=BMU 102;
RX MEDLINE=89345117; PubMed=2503815;
RA Sperka S., Zehelein E., Fiedler S., Sommer R., Buckel P.;
RT "Complete nucleotide sequence of Pseudomonas fluorescens D-galactose
RT dehydrogenase gene.";
RL Nucleic Acids Res. 17:5402-5402(1989).
CC -1- CATALYTIC ACTIVITY: D-galactose + NAD(+) = D-galactono-1,4-lactone
CC + NADH.
CC -1- PATHWAY: FIRST STEP IN D-GALACTOSE CATABOLISM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14913; CAA33038.1; --
DR PIR; S04853; S04853.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR InterPro; IPR004104; GFO_IDH_MOCA_C.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
KW Oxidoreductase; NAD; Galactose metabolism.
SQ SEQUENCE 304 AA; 33049 MW; D842A776DC67A3B0 CRC64;

Alignment Scores:
Pred. No.: 7.88 Length: 304
Score: 76.00 Matches: 35
Percent Similarity: 39.43% Conservative: 34
Best Local Similarity: 20.00% Mismatches: 58
Query Match: 6.93% Indels: 48
DB: 1 Gaps: 6

US-09-914-191-1 (1-598) x GAL_PSEFL (1-304)
QY 10 TTCTTCTTTTATAAAATAGTACTGCGATTAAAAAAGACACTTCTGCCAAGGAACCA 69
DB 73 PheAlaLeuValGlnGlnAlaLeuAlaGlyLysHisValLeuValGluLysProPro 92
QY 70 TGTTCACACACCACCAAGAGTGTCTGCTTAAACAGAGTAAGATACACACCCCATC 129
DB 93 CysAlaThrLeuGlyLysAlaAlaLeuTrpIleLysArgGluGlnAlaSerAlaProCys 112
QY 130 CATCCCTTCCTCCCTGTTCCCTCCCACTTGAGTTGTGTGTCATTCGCACAGTGTCTG 189
DB 113 SerProCysIleAlaTyrAlaPro-----AlaIleAla 123
QY 190 GGTGGTAGGATGCTACAGCCACTAAGCAAGGAGCCCTGGAGGTGGGAGGCTTGA 249
DB 124 AlaAlaArgasp-----127
QY 250 TGTGTAAGCACACACAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTAGAATCTCTG 309
DB 128 TrpLeuAlaThrArgThr-----LeuGlnSerVal 137
QY 310 GATGTTCTTCCAGAAAGATCCCGATGATATCGCAGTGCAGGAGGACTGGCTTGTTC 369
DB 138 GlnIleAspTrpLysGluAspValArgLysTrpHisProGlyGlnAla-TripIleTrpG1 157
QY 370 TGTGCGGGTCACTGCCATCTTTTCTTCATTTCTGTTGGCAGCTTAATTTCTTTG 429
DB 157 n---ProGly-LeuGlyValPheAspProGlyIleAsnAlaLeuSerIleValThr----174
QY 430 TCATCACTTCACTCCACTTCTGCGCATATCAACACAGTCCCTTTCTTATATACATCGCAGCT 489
DB 175 -----HisLeuLeuProLeuPro----LeuPheValGluSerA 186
QY 490 CATTATTATAGTTGATGTTGAATTCAGAAACAAAATCTCA 530
```

```
DR EMBL; AF041335; AAC32805.1; --
* DR EMBL; AF041333; AAC32805.1; JOINED.
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 9, 2003, 12:45:58 ; Search time 47.5 Seconds

(without alignments)
5188.049 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 1096

Sequence: 1 ttggaatagttcttctgctta.....gggttagtcagattgttg 598

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US0914191/runat_09072003.112349.12876/app_query.fasta.1.775
-DB=SPTRMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0914191 @CGN 1.1 138 @runat_09072003.112349.12876 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
c 1	544	49.2	173 4 Q9H8R6	Q9H8R6 homo sapien

c 2	544	49.2	298 4	Q9H0S2	Q9h0s2 homo sapien
c 3	537	48.6	298 4	Q9NWX6	Q9nwx6 homo sapien
c 4	428	38.7	238 11	Q9CQT0	Q9cqt0 mus musculus
c 5	428	38.7	298 11	Q9CY52	Q9cy52 mus musculus
c 6	219	19.8	286 5	Q9V3N8	Q9v3n8 drosophila
c 7	169.5	15.3	261 3	Q9Y7T3	Q9y7t3 schizosacch
c 8	129.5	11.7	495 10	Q9SIQ0	Q9siq0 arabidopsis
c 9	94	8.5	297 10	Q9ZV61	Q9zv61 arabidopsis
c 10	88.5	8.1	457 10	Q93ZL5	Q93zl5 arabidopsis
c 11	88.5	8.1	515 10	Q9FK95	Q9fk95 arabidopsis
c 12	85	7.7	330 12	Q89277	Q89277 bombyx mori
c 13	84.5	7.7	495 11	Q921R1	Q921r1 mus musculus
c 14	82.5	7.5	343 10	Q9LP79	Q9lp79 arabidopsis
c 15	82	7.5	1628 4	Q9HCL7	Q9hcl7 homo sapien
c 16	81.5	7.4	370 10	Q22742	Q22742 arabidopsis
c 17	81.5	7.4	661 10	Q9FVC8	Q9fvc8 arabidopsis
c 18	81.5	7.4	2406 4	Q9BZS0	Q9bzs0 homo sapien
c 19	81	7.3	390 12	Q89276	Q89276 bombyx mori
c 20	80.5	7.3	410 16	Q8YSZ6	Q8ysz6 anabaena sp
c 21	80	7.2	304 3	Q9C1J4	Q9c1j4 pichia past
c 22	80	7.3	402 4	Q9NS06	Q9ns06 homo sapien
c 23	80	7.3	403 4	Q9C009	Q9c009 homo sapien
c 24	80	7.3	484 4	Q9BZ04	Q9bz04 homo sapien
c 25	79.5	7.3	2144 11	Q9Q1P2	Q9q1p2 rattus norv
c 26	79	7.2	359 10	Q9AVE6	Q9ave6 oryza sativ
c 27	78.5	7.2	225 5	Q9NAF1	Q9naf1 caenorhabdi
c 28	78.5	7.2	397 16	Q9K715	Q9k715 bacillus ha
c 29	78	7.1	252 17	Q8TZ46	Q8tz46 methanopyru
c 30	78	7.1	672 11	Q9WVH4	Q9wvh4 mus musculu
c 31	78	7.1	2906 11	Q9WUH9	Q9wuh9 rattus norv
c 32	77	7.0	181 4	Q9UPC2	Q9upc2 homo sapien
c 33	77	7.0	560 4	Q96AW0	Q96aw0 homo sapien
c 34	77	7.0	683 11	Q9DSW6	Q9dsw6 mus musculu
c 35	76.5	6.9	284 16	Q84478	Q84478 chlamydia t
c 36	76.5	6.9	284 16	Q9PJS1	Q9pjs1 chlamydia m
c 37	76.5	7.0	293 3	Q12393	Q12393 saccharomyc
c 38	76.5	6.9	439 2	Q60021	Q60021 thermus aqu
c 39	76.5	6.9	612 10	Q9FGW5	Q9fgw5 arabidopsis
c 40	76.5	6.9	644 15	Q85735	Q85735 murine type
c 41	76.5	6.9	1006 3	Q9C2A4	Q9c2a4 neurospora
c 42	76.5	6.9	1037 16	Q9KRG9	Q9kr99 vibrio chol
c 43	76.5	7.0	1773 10	Q94HV6	Q94hv6 oryza sativ
c 44	76	6.9	268 4	Q9UJD7	Q9ujd7 homo sapien
c 45	76	6.9	327 16	Q8Z0H4	Q8z0h4 anabaena sp

ALIGNMENTS

RESULT 1

Q9H8R6	PRELIMINARY;	PRT;	173 AA.
ID	Q9H8R6		
AC	Q9H8R6;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	CNNA FLJ13293 fis, clone OVARC1001188 (Similar to hypothetical protein		
DE	FLJ20546) (Hypothetical 20.2 kDa protein).		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OVARY;		
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		

RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MUSCLE;
 RA Strauberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]

RN SEQUENCE FROM N.A.
 RP TISSUE=MUSCLE;
 RC Strauberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK023355; BAB14540.1; -;
 DR EMBL; BC001852; AAH01852.1; -;
 DR EMBL; BC001523; AAH01523.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 173 AA; 20157 MW; 70C1D8FFC116B416 CRC64;

Alignment Scores:
 Pred. No.: 3,6e-52 Length: 173
 Score: 544.00 Matches: 103
 Percent Similarity: 99.04% Conservatives: 0
 Best Local Similarity: 99.04% Mismatches: 1
 Query Match: 49.23% Indels: 0
 DB: 4 Gaps: 0

US-09-914-191-1 (1-598) x Q9H8R6 (1-173)

QY 598 CAACAATCTGGACTTAACACCAAGCCAGGAGATTACAGGAACTCTTNCAGCA 539
 |||||
 Db 70 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaA 89
 |||||
 QY 538 GACAAGAATGAGATTGTTTCTGAAATTCAAACATCAACTATATAATGAGTCCCGATG 479
 |||||
 Db 90 AspLysAsnGluLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluLeuPromet 109
 |||||
 QY 478 TATAGAAAGGACTGTGTTGATATATGCGAAGAGTGGATGATGATGATGATGATGATGAT 419
 |||||
 Db 110 TyrArgLysGlyThrValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 129
 |||||
 QY 418 AAGCTGCCAACAGAAATGGAGGAAAGAAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 359
 |||||
 Db 130 LysLeuProThrGluMetGluGlyLysValMetAlaValThrArgThrArgThrLysPro 149
 |||||
 QY 358 GTGCCCTTGCATCGCATATCATCGGGATGCTTCTTGGAGGAACATCCAGAGATTCTA 299
 |||||
 Db 150 ValProLeuHisCysAspIlelleGlyAspAlaPheThrLysGluHisProGluLeuLeu 169
 |||||
 QY 298 GATGAACACAGC 287
 |||||
 Db 170 AspGluAspSer 173

RESULT 2

Q9H0S2 ID Q9H0S2 PRELIMINARY; PRT; 298 AA.
 AC Q9H0S2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical 34.8 kDa protein.
 GN DKF2P564B1172.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=BRIN;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansoerge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehner K., Strack N.,
 RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 *RT "Towards a Catalog of Human Genes and Proteins: Sequencing and

RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
 RL Genome Res. 11:422-435 (2001).
 DR EMBL; AL136669; CAB66604.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 298 AA; 34831 MW; A6C941B75611C448 CRC64;

Alignment Scores:
 Pred. No.: 3,84e-52 Length: 298
 Score: 544.00 Matches: 103
 Percent Similarity: 99.04% Conservatives: 0
 Best Local Similarity: 99.04% Mismatches: 1
 Query Match: 49.23% Indels: 0
 DB: 4 Gaps: 0

US-09-914-191-1 (1-598) x Q9H0S2 (1-298)

QY 598 CAACAATCTGGACTTAACACCAAGCCAGGAGATTACAGGAACTCTTNCAGCA 539
 |||||
 Db 195 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAlaA 214
 |||||
 QY 538 GACAAGAATGAGATTGTTTCTGAAATTCAAACATCAACTATATAATGAGTCCCGATG 479
 |||||
 Db 215 AspLysAsnGluLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluLeuPromet 234
 |||||
 QY 478 TATAGAAAGGACTGTGTTGATATATGCGAAGAGTGGATGATGATGATGATGATGATGAT 419
 |||||
 Db 235 TyrArgLysGlyThrValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 254
 |||||
 QY 418 AAGCTGCCAACAGAAATGGAGGAAAGAAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 359
 |||||
 Db 255 LysLeuProThrGluMetGluGlyLysValMetAlaValThrArgThrArgThrLysPro 274
 |||||
 QY 358 GTGCCCTTGCATCGCATATCATCGGGATGCTTCTTGGAGGAACATCCAGAGATTCTA 299
 |||||
 Db 275 ValProLeuHisCysAspIlelleGlyAspAlaPheThrLysGluHisProGluLeuLeu 294
 |||||
 QY 298 GATGAACACAGC 287
 |||||
 Db 295 AspGluAspSer 298

RESULT 3

Q9NWX6 ID Q9NWX6 PRELIMINARY; PRT; 298 AA.
 AC Q9NWX6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hypothetical 34.8 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=EMBRYO;
 RC Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000553; BAA91249.1; -;
 DR EMBL; AK021663; BAB13870.1; -;

SQ SEQUENCE 298 AA; 34815 MW; A6C951E716019408 CRC64;

Alignment Scores:

Pred. No.: 2,34e-51 Length: 298
 Score: 537.00 Matches: 102
 Percent Similarity: 98.08% Conservatives: 0
 Best Local Similarity: 98.08% Mismatches: 2
 Query Match: 48.60% Indels: 0
 DB: 4 Gaps: 0

US-09-914-191-1 (1-598) x Q9NCX6 (1-298)

QY 598 CAACAATCTGGACTAACACAGTACAGCCCAAGGAGATACAGGGAATCTTTNCAGCA 539
 DB 195 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAla 214
 QY 538 GACAAGAAATGAGATTTTCTGTAATCAACATCAACTATATATGATGAGCTGCCGATG 479
 DB 215 AspLysAsnGluLeuPheSerGluPheAsnIleAsnTyrAsnAsnGluProMet 234
 QY 478 TATAGGAAGGAGACTGTGTTGATATGCGAGAAGTGGATCAAGTGAAGTGAACAAAGAAATT 419
 DB 235 TyrArgLysGlyThrValLeuLeuIleTyrGlnLysValAspGluValMetThrLysGluLe 254
 QY 418 AAGCTGCCAACAGAAATGGAAGAAAGAGATGCGAGTACCCGACGACGACCAAGCCCA 359
 DB 255 LysLeuProThrGluMetGluLysLysMetAlaValThrArgThrArgThrLysPro 274
 QY 358 GTCCCTTCACATGCGATATCATCGGGATGCTTCTCGAAGGACATCCAGAGATCTTA 299
 DB 275 ValProLeuHisCysAspIleLeuGlyAspAlaPheThrLysGluHisProGluLeuLeu 294
 QY 298 GATGAAGACAGC 287
 DB 295 AspGluAspSer 298

RESULT 4

Q9NCX6 PRELIMINARY; PRT; 238 AA.
 AC Q9COT0
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 5730409G07Rik protein.
 GN 5730409G07Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).

DR ENBL; AK016076; BAB30104.1; -;
 DR ENBL; AK015467; BAB29859.1; -;
 DR MGD; MGI:1913878; 5730409G07Rik.
 SQ SEQUENCE 238 AA; 27998 MW; 5EBACDE190A0C42D CRC64;

Alignment Scores:

Pred. No.: 3,95e-39 Length: 238
 Score: 428.00 Matches: 80
 Percent Similarity: 89.42% Conservatives: 13
 Best Local Similarity: 76.92% Mismatches: 11
 Query Match: 38.73% Indels: 0
 DB: 11 Gaps: 0

US-09-914-191-1 (1-598) x Q9COT0 (1-238)

QY 598 CAACAATCTGGACTAACACAGTACAGCCCAAGGAGATACAGGGAATCTTTNCAGCA 539
 DB 135 GlnGlnSerGlyLeuThrProValGlnAlaGlnGlyArgLeuGlnGlyThrLeuAla 154
 QY 538 GACAAGAAATGAGATTTTCTGTAATCAACATCAACTATATATGATGAGCTGCCGATG 479
 DB 155 AspLysAsnGluLeuPheSerGluPheHisIleAsnTyrAsnAsnGluProHisMet 174
 QY 478 TATAGGAAGGAGACTGTGTTGATATGCGAGAAGTGGATCAAGTGAAGTGAACAAAGAAATT 419
 DB 175 TyrArgLysGlyThrValLeuValTyrGlnLysValGluValArgThrGlnGluVal 194
 QY 418 AAGCTGCCAACAGAAATGGAAGAAAGAGTGGCAGTGCACCCGACGACGACCAAGCCCA 359
 DB 195 ValProLeuHisCysAspIleLeuGlyAspAlaPheThrLysGluHisProGluLeuLeu 234
 QY 298 GATGAAGACAGC 287
 DB 235 AlaGluGluAsn 238

RESULT 5

Q9CYS2 PRELIMINARY; PRT; 298 AA.
 AC Q9CYS2
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 5730409G07Rik protein.
 GN 5730409G07Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,

RA Hayaishizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK010876; BAB27240.1; -;
 DR MGI; MGI:1913878; 573040907Rik.
 SQ SEQUENCE 298 AA; 34971 MW; D4CA0C154E452ECA CRC64;

Alignment Scores:
 Pred. No.: 4.06e-39 Length: 298
 Score: 428.00 Matches: 80
 Percent Similarity: 89.42% Conservatives: 13
 Best Local Similarity: 76.92% Mismatches: 11
 Query Match: 38.73% Indels: 0
 DB: 11 Gaps: 0

US-09-914-191-1 (1-598) x 09CY52 (1-298)

QY 598 CAACAATCTGGACTTAACACCAAGGAGGAGATACAGGAACTCTTNCAGCA 539
 |||||
 DB 195 GlnGlnSerGlyLeuThrProValGlnAlaGlnArgLeuLysGlyThrLeuThrAla 214
 |||||

QY 538 GACAGATGAGATTTGTTCTTGAATCAACATCACTATATATGAGCTGCCGATG 479
 |||||
 DB 215 AspLysAsnGluLeuLeuPheSerGluPheHisLeuAsnTyrAsnAsnGluProHisMet 234
 |||||

QY 478 TATAGGAAGGAGCTGTGTGATATGCGACAGAGTGGATGATGACAAAGAAATT 419
 |||||
 DB 235 TyrArgLysGlyThrValLeuValTrpGlnLysValGluValArgThrGlnGluVal 254
 |||||

QY 418 AAGCTGCCAACAGAAATGGAAGGAAAGATGCGAGTGCACCGGACGAGCAAGGCCA 359
 |||||
 DB 255 ArgLeuProAlaGluMetGluGlyGluLysLysAlaValAlaArgThrArgThrArgVal 274
 |||||

QY 358 GTGCCCTTCACCTGCAGATATCATCGGGATGCTTTCTTGAAGCAATCAGAGATCTA 299
 |||||
 DB 275 ValAlaLeuAsnCysAspLeuLeuLeuGlyAspAlaPheTrpLysGluHisProGluLeuLeu 294
 |||||

QY 298 GATGACAGACAGC 287
 |||||
 DB 295 AlaGluGluAsn 298
 |||||

RESULT 6

ID Q9V3N8 PRELIMINARY; PRT; 286 AA.
 AC Q9V3N8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAR-2001 (TrEMBLrel. 13, Last sequence update)
 DE BG:DS00929.4 protein.
 GN BG:DS00929.4 OR CG4103.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
 RA Celniker S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genetics 153:179-219(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003646; AAF53429.1; -;
 DR EMBL; AE003411; AAF44908.1; -;
 DR FlyBase; FBgn0028925; BG:DS00929.4.
 KW Hypothetical protein.
 SQ SEQUENCE 286 AA; 33072 MW; 9539693DA9C2BFC8 CRC64;

Alignment Scores:
 Pred. No.: 1.18e-15 Length: 286
 Score: 219.00 Matches: 47
 Percent Similarity: 60.61% Conservatives: 13
 Best Local Similarity: 47.47% Mismatches: 21
 Query Match: 19.82% Indels: 18
 DB: 5 Gaps: 2

US-09-914-191-1 (1-598) x Q9V3N8 (1-286)

QY 595 CAATCTGAGTAAACACCAAGGAGGAGATACAGGAACTCTTNCAGCAGAC 536
 |||||
 DB 166 GluLysGlyLeuThrAsnGlnAlaGluAlaLysLeuArgGlyThrPheSerAlaAsp 185
 |||||

QY 535 AAGATGAGATTTCTTCTTGAATCAACTCACTATATATGAGCTGCCGATGAT 476
 |||||
 DB 186 LysAsnGluLeuLeuPheGlnGluPheGlyLeuAsnTyrAsnAsnLeuProAlaMetTyr 205
 |||||

QY 475 AGGAAAGGAGCTGTGTGATATGCGAGAGGTGGTGAAGTATGATGACAAAGAAATTAAG 416
 |||||

Alignment Scores:			
Pred. No.:	4,22e-10	Length:	261
Score:	169.50	Matches:	38
Percent Similarity:	59.41%	Conservative:	22
Best Local Similarity:	37.62%	Mismatches:	32
Query Match:	15.34%	Indels:	9
DB:	3	Gaps:	2
US-09-914-191-1 (1-598) x Q9V7T3 (1-261)			
QY	595	CAATCTGACATAACACCAAGTCAACGCCAAGGAGATTACAGGGA	ACTCTTNCAGCAGAC 536
Db	166	LysGlyGlyPheThrAsnThrGlnAlaGluTyrLeuLysGlyThrValSerAlaGlu	185
QY	535	AGAAATGAGATTTTCTTCTGAATTCAAATCAACTATAATAATCAGCTGCCGATGAT	476
Db	186	LysHisGluIleLeuPheSerLysPheGlyIleAsnTyrAsnPheGluProGluIleTyr	205
QY	475	AGGAAGGAGCTGTGTCATATGTCAGAAAGTGGATGAAGTATGATACAAAAGAAATTAAG	416
Db	206	LysLysGlySerIleTrrIleArgGlyProIleAspGlnGluTrpHisGlnGlnAspLys	225
QY	415	CTGCCAACAGAAATGCGAAGGAAAAAGATGCACTGACCCGGACACAGGACAAAGCCAGTG	356
Db	226	LysPheSerValLysGlnLysLysMetValLeuSer	238
QY	355	CCCTTTGCACTGCGATATCATCGGGGATGCTTCTTGGAAAGGAACATCCA	302
Db	239	IleLeuHisValSerLeuIleAspAspPheTrpThrSerArgProPheLeuGluVal	258
QY	301	CTA 299	
Db	259	Leu 259	
RESULT 8			
Q9SIQ0	PRELIMINARY;	PRT;	495 AA.
AC	Q9SIQ0;		

Score:	88.50	Matches:	28
Percent Similarity:	35.87%	Conservative:	5
Best Local Similarity:	30.43%	Mismatches:	32
Query Match:	8.07%	Indels:	27
DB:	10	Gaps:	3

US-09-914-191-1 (1-598) x Q93ZL5 (1-457)

QY	71	GTTCACACACCGCAACAGGTGTCCTTAAACAGAGTAAGATACACACCCCATCC	130
Db	267	ValProleuAsnGlnThrAsnGluGluAlaArgThrVal-----SerProleuProlys	284
QY	131	ATCCCTTCCTCCCTGTTCCCTCCCAACTGAGTTGTGTCATTCGCACCAAGTGCCTGG	190
Db	285	ValProCysPheProGlyProProProThrTrpProTyAlaTrpAsnGlyValSerTrp	304
QY	191	GTGGTAGGGGATGTACAGCCACCTAAGCCAGG-----AGCCCTGGG	232
Db	305	ThrIleuProPheTyProProProAlaTyTrpSerCysProGlyValSerProGly	324
QY	233	AGGTGGGAGGGCTTGTCATGGTTAAGCACACCCAGAACTGAAGCGCAAAAGGTCAGCTGTC	292
Db	325	AlaTrpAsnSerPhe-----	329
QY	293	TTCATCTAGAATCTCTGGATGTTCTTCCTCAGAAAAGC	328
Db	330	-----ThrTrpMetProGlnProAsnSer	337

RESULT 11

Q9FK95	PRELIMINARY;	PRT;	515 AA.
ID	Q9FK95		
AC	Q9FK95;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Similarity to H-protein promoter binding factor.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RX	MEDLINE=38403884; PubMed=9734815;		
RA	Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,		
RA	Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. VI.		
RT	Sequence features of the regions of 1,367,185 bp covered by 19		
RT	physically assigned P1 and TAC clones.;"		
RL	DNA Res. 5:203-216(1998).		
DR	EMBL; AB012243; BAB08998.1; .		
DR	InterPro; IPR003851; Znf_Dof.		
DR	Pfam; PF02701; zf-Dof; 1.		
SQ	SEQUENCE 515 AA; 56528 MW; E2B5BF1957DC33FF CRC64;		

Alignment Scores:

Pred. No.:	0.569	Length:	515
Score:	88.50	Matches:	28
Percent Similarity:	35.87%	Conservative:	5
Best Local Similarity:	30.43%	Mismatches:	32
Query Match:	8.07%	Indels:	27
DB:	10	Gaps:	3

US-09-914-191-1 (1-598) x Q9FK95 (1-515)

QY	71	GTTCACACACCGCAACAGGTGTCCTTAAACAGAGTAAGATACACACCCCATCC	130
Db	267	ValProleuAsnGlnThrAsnGluGluAlaGlyThrVal-----SerProleuProlys	284
QY	131	ATCCCTTCCTCCCTGTTCCCTCCCAACTGAGTTGTGTCATTCGCACCAAGTGCCTGG	190
Db	285	ValProCysPheProGlyProProProThrTrpProTyAlaTrpAsnGlyValSerTrp	304

QY 191 GTGTAGGATGCTACAGCCACCTAAGCAAGG-----AGCCTGGG 232
DB 305 ThrileuProPheTyrProProAlaTyrTrpSerCysProGlyValSerProGly 324
QY 233 AGTGGAGGGCTTCGATGGTTAAGCACACCAAGAACTGAAGCGCAAAAGGTCAGTGC 292
DB 325 AlaTrpAsnSerPhe----- 329
QY 293 TTCATCTAGAATCTCTCGATGTTCTTCCAGAAAGC 328
DB 330 -----ThrTrpMetProGlnProAsnSer 337
RESULT 12
ID O89277 PRELIMINARY; PRT; 390 AA.
AC O89277;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Non-structural protein.
OS Bombyx mori cytoplasmic polyhedrosis virus (BmCPV).
OC Viruses; dsRNA viruses; Reoviridae; Cypovirus.
OX NCBI_TaxID=31594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=98321164; PubMed=9657006;
RA Hagiwara K., Tomita M., Kobayashi J., Miyajima S., Yoshimura T.;
RT "Nucleotide sequences of Bombyx mori cytoplasmic polyhedrosis virus
segment 8.";
RL Biochem. Biophys. Res. Commun. 247:549-553 (1998).
DR EMBL; AB016437; BAA32043.1; -.
SQ SEQUENCE 390 AA; 44151 MW; B172F1D7E790E2B9 CRC64;

Alignment Scores:
Pred. No.: 1.36 Length: 390
Score: 85.00 Matches: 27
Percent Similarity: 50.00% Conservative: 24
Best Local Similarity: 26.47% Mismatches: 43
Query Match: 7.69% Indels: 8
DB: 12 Gaps: 4
US-09-914-191-1 (1-598) x O89277 (1-390)
QY 577 GTCAAGCCCAAGGAGATTACAGGAACTTNCACGACCAAGATGAGATTTCGTT 518
DB 136 IleGluSerGluGlyGluMetAsnAspThrProGlnAspAspGluAlaGluLysLysAsp 155
QY 517 TCTGAATTCAACTCAACTATAATAAT-----GAGCTGCCGATGTATAGAAAGG 467
DB 156 ValGluIleArgMetAsnTrpSerGluGluIleValGluLeuProLysHisGlnGlu 175
QY 466 ACTGTGTTGATATGGCAGAGGTGGATGAGTATGATGACAAAGAAATTAAGTGCACAA 407
DB 176 AsnValLeuValLeuSerLysProSerMetIleSerGluGluGluLeu---MetProThr 194
QY 406 GAAATGAAGAAAAAGATGGCAGTGCACCGGACAGGACAAAGCCAGTCCCTTCAC 347
DB 195 AspMetGlu-----ValAlaThrProLysValLeuGluProProThrProLeuPro 211
QY 346 TGCATATCATCGGGGATGCTTCTGGAAGGAACATCCAGAGATTCCTAGATGAAGACAGC 287
DB 212 SerProIleIle---ValAlaValSerSerGluSerProGlnValLysGluIleGluArg 230
QY 286 TGACCC 281
DB 231 LeuPro 232

RESULT 13
Q921R1 PRELIMINARY; PRT; 495 AA.
ID Q921R1
AC Q921R1;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ectonucleoside triphosphate diphosphohydrolase 2.
GN ENTPD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011241; AAH11241.1; -.
DR MGD; MGI:1096863; Entpd2.
DR InterPro; IPR000407; GDA1_CD39_NTPase.
DR Pfam; PF01150; GDA1_CD39_1.
DR PROSITE; PS01238; GDA1_CD39_NTPASE; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 495 AA; 54319 MW; A76468A0CBF86AAC CRC64;
Alignment Scores:
Pred. No.: 1.59 Length: 495
Score: 84.50 Matches: 42
Percent Similarity: 37.80% Conservative: 20
Best Local Similarity: 25.61% Mismatches: 69
Query Match: 7.71% Indels: 33
DB: 11 Gaps: 9
US-09-914-191-1 (1-598) x Q921R1 (1-495)
QY 87 CAAGGTGTTCTGCTTAAACAGAGTAAATACACACCCCATCCATCCCTTCCTCCCTG 146
DB 250 GlnArgLeuLeuAlaSerAlaLeuGlnIleHisArgPheHisProCysTrpProLysGly 269
QY 147 TTCCTCCCACTTGGTGTGTCATTGCGACCATGTCCTGGGTAGGATGCTAC 206
DB 270 TyrSerThrGlnValLeuLeuArgGluValTyrGlnSerPro-----CysThr 285
QY 207 AGCCACCTAAGCAAGGAGCCCTGGGAGGTGGGAGGCTTGCAATGTTAAGCACACCA 266
DB 286 MetGlyGlnArgProGlnThrPheAsnSerSerAlaThrValSerLeuSerGlyThrSer 305
QY 267 ACTGAAGCGCAAAAGGTCAGTGTCTTCATCT-----AGAATCTCTGGATGTCCTTC 320
DB 306 AsnAlaAlaLeuCysArgAspLeuValSerGlyLeuPheAsnIleSerCysProPhe 325
QY 321 CAGAAAGCATCCCGATGATATCGCATGCAAGGGCAGTGGCTTCTCTGTCGCGGTC 380
DB 326 -----SerGlnCysSerPheAsnGly----- 332
QY 381 ACTGCCATCTTTTCTTCCATTTCTGTGGCAGCTTAATTTCTTTTGTGCATCACTTCA 440
DB 333 -----ValPheGlnProProValAla---GlyAsnPhelIleAlaPhe----- 345
QY 441 TCCACCTTCTGCCATATCAACAGTCCTTCTTATACATCGGCAGCTCATTTATTATAG 500
DB 346 SerAlaPheTyrTyr-----ThrValAspPheLeu-----LysThrValMetGly 360
QY 501 TTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
DB 361 LeuProValGlyThrLeuLysGlnLeuGlu-----AspAlaThrGluThrThrCysAsn 378
QY 561 CTCCTTGGGCT 572
DB 379 GlnThrTrpAla 382
RESULT 14
Q9LP79 PRELIMINARY; PRT; 343 AA.
ID Q9LP79
AC Q9LP79;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE TIN15.7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altafi H., Bei C., Chin C., Chioi J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC TIN15 from chromosome
 I.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC020889; AAF79694.1; -;
 SQ SEQUENCE 343 AA; 38371 MW; CE0EAC1EA62DB64B CRC64;

Alignment Scores:
 Pred. No.: 2.56 Length: 343
 Score: 82.50 Matches: 45
 Percent Similarity: 33.16% Conservative: 18
 Best Local Similarity: 23.68% Mismatches: 62
 Query Match: 7.53% Indels: 65
 DB: 10 Gaps: 6

US-09-914-191-1 (1-598) x Q9LPT9 (1-343)

QY 85 AACAGGTGTTCTGCTTAAC-----AGAGTAAGATACACCCCATCCAT 132
 Db 3 SerLysAlaLeuGlyLeuProGlyPheLeuProGlyArgPheSerProArgLeu 22
 QY 133 CCCTTCCTCCTGCTTCCTCCCACTTCAGTGTGTCATCGCACGAGTCCCTGGGT 192
 Db 23 LeuSerLeuProGlySerProGlyAlaSerThrPheAlaThrArgHisLysLeuAsp 42
 QY 193 GGTAGGATGCTACAGCCACCTAAGGCAAGAGCCCTGGGAGGTGGAGGGCTTCATGG 252
 Db 43 SerArgGlnThr-----LeuLeuTrp 49
 QY 253 TTAGCACACAGAACCTGACGCAAGAGGTGACGTCTTCATCTAGATCTCTGGAT 312
 Db 50 ---AsnLysProGlnLeuSerArgValArgValAlaCysSerSerGlnSerAspSer 68
 QY 313 GTTCTTCAGAAAGCATCCCGATGATGTCAGTGCAGAGGCA---CTGGCTTTGTCC 369
 Db 69 ArgProGluLysLysGlnSerAspLysSerAsnTyrAlaArgAlaGluLeuPheArgGly 88
 QY 370 TGGTCCGGGTCA----- 381
 Db 89 LysSerGlySerValSerPheAsnGlyLeuThrHisGlnLeuValGluGluSerLysLeu 108

QY 381 ----- 381
 Db 109 ValSerAlaProPheGlnGluGlySerPheLeuTrpValLeuAlaProValVal 128
 QY 382 -----CTGCCATCTTTTTCCTTCCATTTCTGTGGCAGCTTAATTT 423
 Db 129 LeuLeuSerSerLeuLeuLeuProGlnPheLeuSerGlyLeuLeuGluAlaThrPhe 148
 QY 424 CTTTGTGATCATCTTCATCCACTTCTGCATATCAACA-----CAGTCCCTT 471
 Db 149 LysAsnAspThrValAlaGlyArgSerGluLeuValThrSerPheCysPheGluThrVal 168
 QY 472 TCTATATACATCGCAGCTCATTTATTATAGT 501
 Db 169 PheTyrAlaGlyLeuAlaLeuPheLeuSer 178

RESULT 15

Q9HCL7 PRELIMINARY; PRT; 1628 AA.
 AC Q9HCL7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE KIAA1555 protein (Fragment).
 GN KIAA1555.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20450683; PubMed=10997877;
 RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
 RT XVIII. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro.";
 RL DNA Res. 7:273-281(2000).
 DR EMBL; AB046775; BAB13381.1; -;
 DR HSSP; P15822; LBBO.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 2.
 DR SMART; SM00355; Znf_C2H2; 2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
 KW DNA-binding; Metal-binding; Zinc-finger.
 FT NON_TER 1
 SQ SEQUENCE 1628 AA; 174427 MW; 4E23A34360A3D939 CRC64;

Alignment Scores:

Pred. No.: 3.51 Length: 1628
 Score: 82.00 Matches: 43
 Percent Similarity: 31.02% Conservative: 24
 Best Local Similarity: 19.91% Mismatches: 77
 Query Match: 7.48% Indels: 72
 DB: 4 Gaps: 9

US-09-914-191-1 (1-598) x Q9HCL7 (1-1628)

QY 38 TTAATAAAAGACACTTCTGCCAAAGAACCATGTTCCAAACCGCAACAGGTGTTCT 97
 Db 1204 LeuAlaArgLysHisSerLeuThrLysAsnAsp---SerSerProGln---ArgCysSer 1221
 QY 98 GCTTAACAGAGTAAGATACACCCCATCCATCCCTCCTTC----- 142
 Db 1222 ProAlaArgGluProGlnAlaSerAlaProSerProGlyLeuHisValAspProGly 1241
 QY 143 -----CCTGTTCCCTCCCAACTGAGTTGTGTGTCATTCGACCCAGGTGCC--- 187
 Db 1242 ArgGlyMetGlyProLeuProCysGlySerProArgLeuGlnLeuSerProLeuThrLeu 1261
 QY 187 ----- 187

```
Db 1262 CysProLeuGlyArgGluLeuAlaProArgAlaHisValLeuSerLysLeuGluGlyThr 1281
Qy 188 -----TGGGTGGTAGGGATGCTACAG 208
Db 1282 ThrAspProGlyLeuProArgTy-SerProThrArgArgTrpSerProGlyGlnAlaGlu 1301
Qy 209 ---CCACCTAAGGCAAGAGCCCTGGAGGTGGAGGGCTTGCATGGTTAAGCACACCAG 265
Db 1302 SerProProArgSerAlaProProGlyLysTrp---AlaLeuAlaGlyProGlySerPro 1320
Qy 266 AACTGAAGCGCAAAAGGTCAGCTCTTCATCTAGAACTCTCGATGTTCTTCCTCCAGAA 325
Db 1321 SerAlaGlyGluHisGlyProGlyLeuGlyLeuAlaProArgValLeuPheProProAla 1340
Qy 326 AGCATCCCCGATGATATCGCAGTCAGGCGGCACTGGCTTTGT----- 367
Db 1341 ProLeuProHisLysLeuLeuSerArgSerProGluThrCysAlaSerProTrpLysAla 1360
Qy 368 -----CTGGTCCGGGTAC----- 382
Db 1361 GluSerArgSerProSerCysSerProGlyProAlaHisProLeuSerSerArgProPhe 1380
Qy 383 -----TGCCATCTTTTTCCTCCATTCTGT-----TGGCAGCTTAATTTCTTTGT 430
Db 1381 SerAlaLeuHisAspPheHisGlyHisIleLeuAlaArgThrGluGluAsnIlePheSer 1400
Qy 431 CATCACTTCATCCACCTTCTGCCATATCACACAGTCCTTCCTATA 478
Db 1401 HisLeuProLeuHisSerGlnHisLeuThrArgAlaProCysProLeu 1416
```

Search completed: July 9, 2003, 12:50:28
Job time : 53.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 13:00:53 ; Search time 1245 Seconds
(without alignments)
13978.697 Million cell updates/sec

Title: US-09-914-191-1
Perfect score: 598
Sequence: 1 ttggaatagttcttctgttta.....gggttagtcacagattgttg 598

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_pro.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	597	99.8	598	6	AX034339	AX034339 Sequence
2	597	99.8	1208	9	BC001523	BC001523 Homo sapi
3	597	99.8	1208	9	BC001852	BC001852 Homo sapi
4	593.8	99.3	1233	9	AK000553	AK000553 Homo sapi
5	591.2	98.9	2159	9	AK000953	AK000953 Homo sapi
6	590.6	98.8	1216	9	HS0801637	AL136669 Homo sapi
7	577.6	96.6	2263	9	AK021663	AK021663 Homo sapi
8	504	84.3	1780	9	AK023355	AK023355 Homo sapi
9	447	74.7	155344	9	AC026407	AC026407 Homo sapi
10	406.4	68.0	416	9	HUMZD93F03	AF086486 Homo sapi
11	370.8	62.0	374	11	G38490	G38490 SHGC-58349
12	306.6	51.3	567	11	G37344	G37344 SHGC-57583
13	177.4	29.7	207877	10	AL645948	AL645948 Mouse DNA
14	167	27.9	318930	2	AC073495	AC073495 Mus muscu
15	78.4	13.1	202051	2	AC128452	AC128452 Rattus no
16	50.8	8.5	37919	8	SPCC63	AL049522 S.pombe c
17	50.2	8.4	18333	2	AC019777	AC019777 Drosophil
18	50.2	8.4	184682	3	AC092238	AC092238 Drosophil
19	50.2	8.4	266308	3	AE003646	AE003646 Drosophil
20	50.2	8.4	303043	3	DROSADH05	AE003411 Drosophil
21	48.4	8.1	7218	6	I66494	I66494 Sequence 14
22	44.2	7.4	23463	2	AC020869	AC020869 Mus muscu
23	43.8	7.3	1719	8	SCYGR024C	Z72809 S.cerevisia
24	42.2	7.1	125020	9	AF429315	AF429315 Homo sapi
25	41.4	6.9	95746	2	AC097111	AC097111 Oryza sat
26	41.4	6.9	146261	2	AC097112	AC097112 Oryza sat
27	41.2	6.9	94075	2	AC131508	AC131508 Strongylo
28	41	6.9	165327	2	AC120699	AC120699 Rattus no
29	40.8	6.8	160759	2	AC117082	AC117082 Dictyoste
30	40.6	6.8	171115	2	AC102661	AC102661 Mus muscu
31	40	6.7	159693	10	AL646042	AL646042 Mouse DNA
32	40	6.7	224076	2	AC084019	AC084019 Mus muscu
33	39.4	6.6	117432	9	AC108704	AC108704 Homo sapi
34	39.4	6.6	151552	2	AC023437	AC023437 Homo sapi
35	39.4	6.6	227197	2	AC126179	AC126179 Homo sapi
36	39.2	6.6	43795	9	AC000022	AC000022 Genomic s
37	39.2	6.6	85472	9	AC025735	AC025735 Homo sapi
38	39.2	6.6	157706	2	AC073560	AC073560 Homo sapi
39	39.2	6.6	178542	2	AC096064	AC096064 Rattus no
40	39.2	6.6	186838	9	AC006338	AC006338 Homo sapi
41	39	6.5	12940	3	AE001386	AE001386 Plasmodiu
42	38.8	6.5	230487	2	AC116390	AC116390 Mus muscu
43	38.6	6.5	209973	2	AC073755	AC073755 Mus muscu
44	38.4	6.4	10212	3	DDI114911	AJ314911 Dictyoste
45	38.4	6.4	106144	2	AC116425	AC116425 Dictyoste

ALIGNMENTS

RESULT 1	AX034339	Sequence 1	598 bp	DNA	linear	PAT 22-SEP-2000
LOCUS	AX034339	Sequence 1	from Patent WO0050637.			
DEFINITION	AX034339	Sequence 1	from Patent WO0050637.			
ACCESSION	AX034339	Sequence 1	from Patent WO0050637.			
VERSION	AX034339.1	GI:10303095				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 598)					
AUTHORS	Godson,C.M., Brady,H.R. and Martin,F.M.					
TITLE	Identification of genes having a role in the presentation of					
	diabetic nephropathy					

JOURNAL Patent: WO 0050637-A 1 31-AUG-2000;
GODSON CATHERINE MARY (IE); BRADY HUGH REDMOND (IE); HIBERGEN
LIMITED (IE); MARTIN FINIAN MARY (IE); UNIV COLLEGE DUBLIN
NATIONAL U (IE)
FEATURES
source
1. .598
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 145 a 122 g 176 t 1 others
ORIGIN
Query Match 99.8%; Score 597; DB 6; Length 598;
Best Local Similarity 100.0%; Pred. No. 2.6e-160;
Matches 598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TTGGAATAGTCTGCTTTTAAATAATAGTACGCGATTAATAAAAGACACTTCTGCCA 60
Db 1 TTGGAATAGTCTGCTTTTAAATAATAGTACGCGATTAATAAAAGACACTTCTGCCA 60
Qy 61 AAGGAACCATGTTCCACACCGCAACAGGTGTTCTGCTTAAACAGAGTAAGATACACC 120
Db 61 AAGGAACCATGTTCCACACCGCAACAGGTGTTCTGCTTAAACAGAGTAAGATACACC 120
Qy 121 ACCCCCATCATCCCTTCT 180
Db 121 ACCCCCATCATCCCTTCT 180
Qy 181 AGTGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGG 240
Db 181 AGTGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGG 240
Qy 241 GGGCTTGCATGTTAGCACACAGCACTGAAGCGCAAGAGGTGAGCTTCTTCTATCA 300
Db 241 GGGCTTGCATGTTAGCACACAGCACTGAAGCGCAAGAGGTGAGCTTCTTCTATCA 300
Qy 301 GAATCTCTGATGTTCTTCCAGAAAGCATCCCGGATGATATCGCAGTGAAGGGCACTG 360
Db 301 GAATCTCTGATGTTCTTCCAGAAAGCATCCCGGATGATATCGCAGTGAAGGGCACTG 360
Qy 361 GCTTTGCTCGTCCGGTCACTGCCATCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 361 GCTTTGCTCGTCCGGTCACTGCCATCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Qy 421 TTTCTTTTGCATCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 480
Db 421 TTTCTTTTGCATCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 480
Qy 481 TCGGCAGCTCATTTATATAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTG 540
Db 481 TCGGCAGCTCATTTATATAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTG 540
Qy 541 CTGNAAGAGTTCCTGTAAATCTCCCTGGGCTTGTACTGTTAGTCCAGATTGTTG 598
Db 541 CTGNAAGAGTTCCTGTAAATCTCCCTGGGCTTGTACTGTTAGTCCAGATTGTTG 598
RESULT 2
BC001523/c
LOCUS 1208 bp mRNA linear PRI 22-OCT-2001
DEFINITION Homo sapiens, clone MGC:2901 IMAGE:3010614, mRNA, complete cds.
ACCESSION BC001523
VERSION BC001523.1 GI:16306697
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1208)
Strausberg R.
Direct Submission
Submitted (21-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 3 Row: b Column: 4.
FEATURES
source
1. .1208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:2901 IMAGE:3010614"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH MGC 17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
383. .904
/codon_start=1
/product="Unknown (protein for MGC:2901)"
/protein_id="AAH01523.1"
/db_xref="GI:16306698"
/translation="MTHVASOPASSYFVWRDYPEDOLLYPPGDFGVVYPSNOTL
KDYLSWRQADCHNNLYNTVFWALIQOGLTPVOAGRLQGTLAADKNEILSEFNIN
YNELPMTKRGTVLIVQKQDEVMTKEIKLPTMEGKRWAVTRTKPVLHCDIIGDA
FWKEHPETLDEDS"
BASE COUNT 346 a 258 c 315 g 289 t
ORIGIN
Query Match 99.8%; Score 597; DB 9; Length 1208;
Best Local Similarity 99.8%; Pred. No. 2.7e-160;
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TTGGAATAGTCTTCTCTTTTATAAAATAGTACTGCGATTAAAAAGCACTTCTGCCA 60
Db 1187 TTGGAATAGTCTTCTCTTTTATAAAATAGTACTGCGATTAAAAAGCACTTCTGCCA 1128
Qy 61 AAGGAACCATGTTCCAAACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACC 120
Db 1127 AAGGAACCATGTTCCAAACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACC 1068
Qy 121 ACCCCCATCCATCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
Db 1067 ACCCCCATCCATCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1008
Qy 181 AGTGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGG 240
Db 1007 AGTGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGG 948
Qy 241 GGGCTTGCATGTTAAGCACACCACTGAAGCGCAAGAGGTGAGTGTCTTCTATCTA 300
Db 947 GGGCTTGCATGTTAAGCACACCACTGAAGCGCAAGAGGTGAGTGTCTTCTATCTA 888
Qy 301 GAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTG 360
Db 887 GAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTG 828
Qy 361 GCTTTGCTCTGCTCGGGTCACTGCCATCTTTTCTTCTCTCTCTCTCTCTCTCTCTCT 420
Db 827 GCTTTGCTCTGCTCGGGTCACTGCCATCTTTTCTTCTCTCTCTCTCTCTCTCTCTCT 768
Qy 421 TTTCTTTTGTCTATCATCTTCACTCCACTTCTGCGCATATCAACACAGTCCCTTCTCTATA 480
Db 767 TTTCTTTTGTCTATCATCTTCACTCCACTTCTGCGCATATCAACACAGTCCCTTCTCTATA 708


```

QY 481 TCGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAACAAAATCTCTTCTGTCTG 540
Db 707 TCGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAACAAAATCTCTTCTGTCTG 648
QY 541 CTGNAAGAGTTCCTGTAATCTCCCTGGGCTGTACTGGTGTAGTCCAGATTGTTG 598
Db 647 CTGCAAGAGTTCCTGTAATCTCCCTGGGCTGTACTGGTGTAGTCCAGATTGTTG 590

RESULT 3
BC001852/c
LOCUS BC001852 1208 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to hypothetical protein FLJ20546, clone
ACCESSION BC001852
VERSION BC001852.1 GI:12804814
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Strausberg R.
AUTHORS Direct Submission
TITLE Submitted (29-JAN-2001) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 10 Row: g Column: 20.
Location/Qualifiers
1. .1208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:4529 IMAGE:3010614"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH MGC 17"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
383..904
/codon_start=1
/product="Similar to hypothetical protein FLJ20546"
/protein_id="AAH01852.1"
/db_xref="GI:12804815"
/translation="MPTVASQFASVYVWRDYFEDQPLLYPPGDFRVVVPNOTL
KYLRSRQADCHNNLYNVFWALIQSGLTVPQAGRLQGLTAAKNNLLFSEFN
YNKELPMYRGTVLLIWQKVDVEMTKIKLPTMEGKMAVTRTRTKVPLHCLDIIGA
FWKHEPEILDEDS"

BASE COUNT 346 a 258 c 315 g 289 t
ORIGIN

Query Match 99.8%; Score 597; DB 9; Length 1208;
Best Local Similarity 99.8%; Pred. No. 2.7e-160;
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGGAATAGTCTTGCTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
Db 1187 TTGGAATAGTCTTGCTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 1128
QY 61 AAGGAACCATGTTCCAAACCGCAACAGGTTGTTCTGTTAAACAGAGTAAGATACACC 120

```

```

Db 1127 AAGGAACCATGTTCCAAACCGCAACAGGTTGTTCTGTTAAACAGAGTAAGATACACC 1068
QY 121 ACCCCCATCATCCCTTCCCTTCCCTTCCCTTCCCAACTTGAAGTTGTGTCATTTCCGACC 180
Db 1067 ACCCCCATCATCCCTTCCCTTCCCTTCCCTTCCCAACTTGAAGTTGTGTCATTTCCGACC 1008
QY 181 AGTGTCTCGGTGGTAGGATGCTACAGCCACTAAGGCAAGAGAGCCCTGGGAGGTGGGA 240
Db 1007 AGTGTCTCGGTGGTAGGATGCTACAGCCACTAAGGCAAGAGAGCCCTGGGAGGTGGGA 948
QY 241 GGGCTTGCATGTTAAGCACACAGCACTGAAGCCGCAAAAGGTCAGCTGCTTTCATCTA 300
Db 947 GGGCTTGCATGTTAAGCACACAGCACTGAAGCCGCAAAAGGTCAGCTGCTTTCATCTA 888
QY 301 GAATCTCTGGATGTTCTCTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGGGCACTG 360
Db 887 GAATCTCTGGATGTTCTCTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGGGCACTG 828
QY 361 GCTTTGCTCTGGTCCGGGTCACTGCCATCTTTTCTTCCATTTCTGTGGCAGCTTAA 420
Db 827 GCTTTGCTCTGGTCCGGGTCACTGCCATCTTTTCTTCCATTTCTGTGGCAGCTTAA 768
QY 421 TTTCCTTTGTCATCACTTCACTCCACCTTCTGCCATATCAACAGTCCCTTTCCTATACA 480
Db 767 TTTCCTTTGTCATCACTTCACTCCACCTTCTGCCATATCAACAGTCCCTTTCCTATACA 708
QY 481 TCGGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAACAAAATCTCTTCTGCTG 540
Db 707 TCGGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAACAAAATCTCTTCTGCTG 648
QY 541 CTGNAAGAGTTCCTGTAATCTCCCTGGGCTGTACTGGTGTAGTCCAGATTGTTG 598
Db 647 CTGNAAGAGTTCCTGTAATCTCCCTGGGCTGTACTGGTGTAGTCCAGATTGTTG 590

RESULT 4
AK000553 1233 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens CDNA FLJ20546 fis, clone KAT11504.
DEFINITION AK000553
ACCESSION AK000553.1 GI:7020725
VERSION oligo capping; fis (full insert sequence).
KEYWORDS Homo sapiens signet-ring cell carcinoma cell_line:KATO III cDNA to
SOURCE mRNA, clone lib:KAT clone:KAT11504.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
TITLE UNpublished
JOURNAL 2 (bases 1 to 1233)
REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
AUTHORS Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
JOURNAL University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. .1233
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


QY 482 CGGAGCTCATTTATAGTGTGATGTTGAATTCAGAAAAAATAATCTCATCTTGTCTGC 541
 Db 1307 CGGCGGCTCATTTATAGTGTGATGTTGAATTCAGAAAAAATAATCTCATCTTGTCTGC 1248
 QY 542 TGAAGAGTTCCTGTAACTCCCTCGGGCTTGACTGGTGTAGTCCAGATTGTTG 598
 Db 1247 TGAAGAGTTCCTGTAACTCCCTCGGGCTTGACTGGTGTAGTCCAGATTGTTG 1191

RESULT 6

HSW801637/c 1216 bp mRNA linear PRI 10-MAR-2001
 LOCUS Homo sapiens mRNA; cDNA DKFZp564B1172 (from clone DKFZp564B1172);
 DEFINITION complete cds.

ACCESSION AL136669
 VERSION AL136669.1 GI:12052861

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1216)
 Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,

AUTHORS

Ansoerge W., Boescher M., Bloeker H., Bauersachs S., Blum H.,
 Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
 Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
 Wambutt R., Korn B., Klein M., and Poustka A.

TITLE

Toward a Catalog of Human Genes and Proteins: Sequencing and

JOURNAL

Analysis of 500 Novel Complete Protein Coding Human cDNAs

PUBMED

Genome Res. 11 (3), 422-435 (2001)

REFERENCE

1 (bases 1 to 1216)
 Blum H., Bauersachs S., Mewes H.W., Gassenhuber J. and Wiemann S.

AUTHORS

Submitted (15-DEC-2000) MIPS, Am Klopferspitze 18a, D-89152
 Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by LMU (Ludwig Maximilians University,
 Munich/Germany) within the cDNA sequencing consortium of the German
 Genome Project.

COMMENT

This clone (DKFZp564B1172) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

source

1. .1216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="5"
 /clone="DKFZp564B1172"
 /tissue type="brain"
 /clone_lib="564 (synonym: hfr2) . Vector pAMP1; host
 XL-2blue; sites NotI + SalI"
 /dev stage="fetal"
 21. -917
 /gene="DKFZp564B1172"
 21. .917
 /gene="DKFZp564B1172"
 /note="FLJ20546 fis"
 /codon start=1
 /product="hypothetical protein"
 /protein_id="CA866604.1"
 /db_xref="GI:12052862"

gene

CDS

/translations="MWGACKVQVHDSLATISITLRLYLRLGATMAKSFYVDFRAD
 QDCLACWVVRDLGRHFRFAEKHNFAPKENDRALQMTKCAQTWMEELDIYAG
 TSDFGKVFVRKTNFKFRASFKWTHVASQFASVYFWRDYFEDDPLYPGFGDRV
 VVPSNLTLDKXLSWQADCHINNLYNFWALIQOGLTFVQAGRLQGLTAAADKNE
 ILFSEFNINYNLPLWRYKGTVLINQWQVDEWVTKIKLPTMEGKKMAYTRTKFVP
 LHCDIIGDAFWKHEPILDEDS"
 1193

polyA_site

BASE COUNT 342 a 261 c 319 g 294 t

ORIGIN

Query Match 98.8%; Score 590.6; DB 9; Length 1216;
 Best Local Similarity 99.2%; Pred. No. 1.9e-158;
 Matches 593; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTGGAATAGTCTTGTGCTTTATAAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60

Db 1200 TTTTGATAGTCTTGTGCTTTATAAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 1141

QY 61 AAGGAACCATGTTCCAAACACCGCAACAGGTGTTCTGTTTAAACAGAGTAACATACACC 120

Db 1140 AAGGAACCATGTTCCAAACACCGCAACAGGTGTTCTGTTTAAACAGAGTAACATACACC 1081

QY 121 ACCCCCATCCATCCCTTCCCTGTTCCCTGTTCCCTCCCACTTGAAGTGTGTCAATTCGCACC 180

Db 1080 ACCCCCATCCATCCCTTCCCTGTTCCCTGTTCCCTCCCACTTGAAGTGTGTCAATTCGCACC 1021

QY 181 AGTGTCTGGGTGTTAGGATGCTACAGCCACCTAAGGCAAGGAGCCCTGGAGGTGGGA 240

Db 1020 AGTGTCTGGGTGTTAGGATGCTACAGCCACCTAAGGCAAGGAGCCCTGGAGGTGGGA 961

QY 241 GGGCTTGCATGGTTAAGCACACACAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTA 300

Db 960 GGGCTTGCATGGTTAAGCACACACAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTA 901

QY 301 GAATCTCTGGATGTTCCCTTCCAGAAAGCATCCCGCATGATATCGCAGTGCAGGGCACTG 360

Db 900 GAATCTCTGGATGTTCCCTTCCAGAAAGCATCCCGCATGATATCGCAGTGCAGGGCACTG 841

QY 361 GCTTTGCTCTGGTCCGGTCACTGCCATCTTTTTCCTCCCATCTTCTTGTGGCAGCTTAA 420

Db 840 GCTTTGCTCTGGTCCGGTCACTGCCATCTTTTTCCTCCCATCTTCTTGTGGCAGCTTAA 781

QY 421 TTTCTTTTCTCATCACTTCATCCACCTTCGCCATATCAACAGTCCCTTTCCTTATACA 480

Db 780 TTTCTTTTCTCATCACTTCATCCACCTTCGCCATATCAACAGTCCCTTTCCTTATACA 721

QY 481 TCGGCAGCTCATTTATATAGTGTGATGTTGAATTCAGAAAAAATAATCTCATCTTGTCTG 540

Db 720 TCGGCAGCTCATTTATATAGTGTGATGTTGAATTCAGAAAAAATAATCTCATCTTGTCTG 661

QY 541 CTGNAAGAGTTCCTGTAAATCTCCCTTGGGCTTGACTGGTGTAGTCCAGATTGTTG 598

Db 660 CTGNAAGAGTTCCTGTAAATCTCCCTTGGGCTTGACTGGTGTAGTCCAGATTGTTG 603

RESULT 7

AK021663/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5' - & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
1. .2263
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003893"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/notes="cloning vector: pME183FL3"
4. .900
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB13870.1"
/db_xref="GI:10432894"
DTCLAHCVVRLDGRNFRFAEKHNFAPNDSRALQMLKCAOTVMELEDIVIAYG
translation="MVGACKVKVHDSLATISITRLRYLRIGATMAKSKPEYVRDFEAD
QSEYSEFVKRTNWKBRASKEWTHVAFQFASVYVYWRDYFEDQPLLPPGFGRV
VYVPSNQLKYLSWQADCHNNLYNFWALIQOSGLTPVQAGRLGTLAARDKE
ILPSEFNINNPFPYRGVTLQWQVDEWMTKEIKLPTEMEGKKMAVTRTRTRKVP
LHCDDIIGDAFWEKHPHILDEDS"
BASE COUNT 624 a 486 c 580 g 573 t
ORIGIN

Query Match 96.8%; Score 577.6; DB 9; Length 2263;
Best Local Similarity 98.8%; Pred. No. 1e-154;
Matches 592; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 2 TGGAAATAGTCTTGGCTTTATAAAATAGTACTCGATTAAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 61
DB 1184 TGGAAATAGTCTTGGCTTTATAAAATAGTACTCGATTAAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 1125
QY 62 AGGAACCATGTTCCAAACACCGCAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 121
DB 1124 AGGAACCATGTTCCAAACACCGCAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 1065
QY 122 CCCCCATCCATCCCTTCTTCCCTGTTCCCTCCACAC--TTGAGTGTGCTATTGCGAC 179
DB 1064 CCCCCATCCATCCCTTCTTCCCTGTTCCCTCCACACATTTTGTGTGCTATTGCGAC 1005
QY 180 CAGTGTCTGGGTGGTGGATGCTCAGGCACCTAAGGCAAGGAGCCCTGGGAGGTGGG 239
DB 1004 CAGTGTCTGGGTGGTGGATGCTCAGGCACCTAAGGCAAGGAGCCCTGGGAGGTGGG 945
QY 240 AGGGCTTGATGTTTAAAGCACACAGAACTGAAGCGCAAAAGGCTCAGTGTCTTCTATCT 299
DB 944 AGGGCTTGATGTTTAAAGCACACAGAACTGAAGCGCAAAAGGCTCAGTGTCTTCTATCT 885
QY 300 AGAATCTCGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGCGACT 359
DB 884 AGAATCTCGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGCGACT 825
QY 360 GGCTTTGCTGCTGGTGGGTCAGTGCATCTTTTCTTCCATTTCTGTTGGCAGCTTA 419
DB 824 GGCTTTGCTGCTGGTGGGTCAGTGCATCTTTTCTTCCATTTCTGTTGGCAGCTTA 765
QY 420 ATTTCTTTTGTGATCACTTCCATCCACTCTGCGATATCAACAGTCCCTTTCTTATAC 479
DB 764 ATTTCTTTTGTGATCACTTCCATCCACTCTGCGATATCAACAGTCCCTTTCTTATAC 705
QY 480 ATCGGAGCTCATATATATAGTTGATGTTGAATTCAGAAAACAAATCTCATTTCTGTCT 539
DB 704 ATCGGAGCTCATATATATAGTTGATGTTGAATTCAGAAAACAAATCTCATTTCTGTCT 645
QY 540 GCTGNAAGAGTTCCTCTGTAATCTCCCTTTGGGCTGTGACTGGTGTAGTCCAGATTTGTTG 598

DB 644 GCTGCAAGAGTTCCTCTGTAATCTCCTTGGGCTTGTAGTGGTGTAGTCCAGATTGTTG 586
RESULT 8
AK023355/c
LOCUS AK023355.1 GI:10435255
DEFINITION Homo sapiens cDNA FLJ13293 fis, clone OVARC1001188.
ACCESSION AK023355
VERSION AK023355.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone lib:OVARC1
clone:OVARC1001188.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai,T., Ota,T., Hayaashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1780)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5' - & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
1. .1780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OVARC1001188"
/tissue_type="ovary, tumor tissue"
/clone_lib="OVARC1"
/note="cloning vector: pME183FL3"
42. .563
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14540.1"
/db_xref="GI:10435256"
/translation="MTHVASQFASVYFWRDYFEDQPLLPPGFGRVVVPSNQL
KDYLSWQADCHNNLYNFWALIQOSGLTEVQAGRLGTLAADKNILPSEFNIN
YNELPMYRKGVTLQWQVDEWMTKEIKLPTEMEGKKMAVTRTRTRKPVLPVHCDIIGDA
FWKEHPHILDEDS"
BASE COUNT 513 a 368 c 435 g 464 t
ORIGIN

Query Match 84.3%; Score 504; DB 9; Length 1780;
Best Local Similarity 93.0%; Pred. No. 1.4e-133;
Matches 555; Conservative 0; Mismatches 1; Indels 41; Gaps 1;

QY 2 TGAATAGTCTTGGCTTTATAAAATAGTACTCGATTAAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 61
DB 804 TGAATAGTCTTGGCTTTATAAAATAGTACTCGATTAAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 745
QY 62 AGGAACCATGTTCCAAACACCGCAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 121
DB 744 AGGAACCATGTTCCAAACACCGCAACCAAGGCTTCTGCTTTAAACAGAGTAAAGTACACCA 685
QY 122 CCCCCATCCATCCCTTCTTCCCTGTTCCCTCCCACTTGAGTTGTGCTTCGACCA 181

SHGC-57583 G37344;

JOURNAL
COMMENT
SUBMITTED
University
submitted

JOURNAL
COMMENT

Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES
source
Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:357053"
/clone_lib="Scars_fetal_heart_NbHH19W"
BASE COUNT 113 a 91 c 109 g 103 t
ORIGIN

Query Match 68.0%; Score 406.4; DB 9; Length 416;
Best Local Similarity 99.8%; Pred. No. 1.4e-105;
Matches 407; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
DB 408 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 349
QY 61 AAGGAACCATGTTCCAAACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 120
DB 348 AAGGAACCATGTTCCAAACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 289
QY 121 ACCCCATCATCCCTTCTTCCCTGTTCCTCCCAACTTGAAGTGTGTCATTCGCACC 180
DB 288 ACCCCATCATCCCTTCTTCCCTGTTCCTCCCAACTTGAAGTGTGTCATTCGCACC 229
QY 181 AGTGCTCTGGTGGTAGGATGTCTACAGCCACTTAAGGCAAGGAGCCCTGGAGGTGGGA 240
DB 228 AGTGCTCTGGTGGTAGGATGTCTACAGCCACTTAAGGCAAGGAGCCCTGGAGGTGGGA 169
QY 241 GGGCTTGCATGTTAAGCACACAGCACTGAAGGCAAAAGGCTCAGCTGTCTTCATCTA 300
DB 168 GGGCTTGCATGTTAAGCACACAGCACTGAAGGCAAAAGGCTCAGCTGTCTTCATCTA 109
QY 301 GAATCTCTGGATGTCTCTCCAGAAAGCATCCCGGATGATATCGCAGTGAAGGGCACTG 360
DB 108 GAATCTCTGGATGTCTCTCCAGAAAGCATCCCGGATGATATCGCAGTGAAGGGCACTG 49
QY 361 GCTTTGCTCTGGTCCGGGTCACTGCCATCTTTTCTTCCATTTCTG 408
DB 48 GCTTTGCTCTGGTCCGGGTCACTGCCATCTTTTCTTCCATTTCTG 1

RESULT 11
G38490
LOCUS
DEFINITION SHGC-58349 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION G38490
VERSION G38490.1 GI:3249264
KEYWORDS STS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 374)
Myers,R.M.
Human STSs (1998)
Unpublished (1998)
COMMENT

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: CAACACCGCAACCAAGGTGT
Primer B: GCTTAACCATGCAAGCCCTC
STS size: 184
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifiaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Prepared with primer pairs derived from W93044 -- Unigene.

FEATURES

source
Location/Qualifiers
1..374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5"
/clone_lib="Human"

STS
primer_bind 75..258
primer_bind complement(239..258)
BASE COUNT 99 a 97 c 87 g 91 t
ORIGIN

Query Match 62.0%; Score 370.8; DB 11; Length 374;
Best Local Similarity 99.5%; Pred. No. 2.3e-95;
Matches 372; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGGAATAGTCTTCTGTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
DB 1 TTGGAATAGTCTTCTGTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
QY 61 AAGGAACCATGTTCCAAACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 120
DB 61 AAGGAACCATGTTCCAAACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 120
QY 121 ACCCCATCATCCCTTCTTCCCTGTTCCTCCCAACTTCAGTGTGTCATTCGCACC 180
DB 121 ACCCCATCATCCCTTCTTCCCTGTTCCTCCCAACTTCAGTGTGTCATTCGCACC 180
QY 181 AGTGCTCTGGTGGTAGGATGTCTACAGCCACTTAAGGCAAGGAGCCCTGGAGGTGGGA 240
DB 181 AGTGCTCTGGTGGTAGGATGTCTACAGCCACTTAAGGCAAGGAGCCCTGGAGGTGGGA 240
QY 241 GGGCTTGCATGTTAAGCACACCACTGAAGCCCAAGGCTCAGCTGTCTTCATCTA 300
DB 241 GGGCTTGCATGTTAAGCACACCACTGAAGCCCAAGGCTCAGCTGTCTTCATCTA 300

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-298M7 is from the RPI-23 Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

This sequence is the entire insert of clone RP23-298M7. The true left end of clone RP23-29H5 is at 164150 in this sequence. The true right end of clone RP23-302J15 is at 69639 in this sequence.

FEATURES

```

source          1..207877
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /chromosome="11"
                 /clone="RP23-298M7"
                 /clone_lib="RPI-23"
                 91735..91776
                 /note="Sequence from uni-directional dGTP big dye
                 terminator reads only."
                 20490..204554
                 /note="Sequence from uni-directional primer reads and dGTP
                 big dye terminator reads only."
                 20483..204903
                 /note="Sequence from overlapping clone RP23-29H5
                 (AL663031). Assembly confirmed by restriction digest."
                 206726..206753
                 /note="Sequence from overlapping clone RP23-29H5
                 (AL663031). Assembly confirmed by restriction digest."
BASE COUNT    57557 a 45850 c 45787 g 58683 t
ORIGIN

```

```

Query Match      29.7%; Score 177.4; DB 10; Length 207877;
Best Local Similarity 72.7%; Pred. No. 1.1e-39;
Matches 271; Conservative 0; Mismatches 96; Indels 6; Gaps 3;

QY 84 AACACAGGTGTTCTGCTTAACAGAGTAAGATACACACCCCATCCATCCCTCTCTCC 143
Db 172211 AAAAAAGGTGTTCTGCTTAAGCAGAACAGATACACACCCCATATATCCC-TCCTTCA 172269

QY 144 CTGTTCCCTCCCAACTGAGTTGTGTCATTCGCACAGTGCTCTGGGTGGTAGGATGC 203
Db 172270 TCTGTCTCCACCATCTCCAGTCATGCTCTCTTACAGTGTCTTGGACACGAGTGC 172329

QY 204 TACAGGCACCTAAGGCAAGAGCCCTCGGAGGTGGAGGCTTGCATGTTAAGCACACC 263
Db 172330 TTGAACCCACTAGGGTTAGGG- - -GAGACTCGGGAGCCTTGTGTGATTAGCATGCT 172385

QY 264 AGAAGTCAAGCGCAAAAGGTCAGCTGTCTTCATCTAGATCTCTGGATGTTCTTCCAG 323
Db 172386 AGAAGTCAAGCGGTGA-GGTCAAGTCTCTTTCGCCAGAACTCTCGGTGTTCTTCCAG 172444

QY 324 AAAGCATCCCGCATGATATCGCAGTCAAGGGCACTGGCTTGTCTGTGTCGGGTCACT 383
Db 172445 AAGCATCCCATATAGATACAGATCAAGGCCACACCCGGTCTTGTTCAGGCACG 172504

QY 384 GCCATCTTTTCTTCCATTTCTGTGGCAGCTTAATTTCTTTTGTATCATCTCATCC 443
Db 172505 GCCTTCTTTTCCCTTCCATTTCTGTGGCAGCTTCTTGTGTGTGACTTCTTCC 172564

QY 444 ACCTTCTGCCATA 456
Db 172565 ACCTATAGACAAA 172577

```

RESULT 14

AC073495

LOCUS

DEFINITION

SEQUENCE, 64 unordered pieces.

AC073495

VERSION
KEYWORDS
SOURCE
ORGANISM

AC073495.12 GI:10801940
HTG; HTGS PHASE1; HTGS_DRAFT.
Mus musculus

REFERENCE

AUTHORS

1. (bases 1 to 318930)
Metzker, M.D., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dedrich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (19-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 15, 2000 this sequence version replaced gi:10800205.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: MACH

Center clone name: RP23-345K4

Chemistry: Dye-terminator Big Dye: 71% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 282577 bases at least Q40

Consensus quality: 299603 bases at least Q30

Estimated insert size: 298164; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 64 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 100449: contig of 100449 bp in length

* 100450: gap of unknown length

* 100550: contig of 43298 bp in length

* 143848: gap of unknown length

* 143948: contig of 8446 bp in length

* 152393: contig of 8446 bp in length

* 152394: gap of unknown length

* 152493: contig of 7933 bp in length

* 160426: gap of unknown length

* 160526: gap of unknown length

* 160643: contig of 6117 bp in length

* 166544: gap of unknown length

* 166744: contig of 6741 bp in length


```
* 173485 173584: gap of unknown length
* 173585 177408: contig of 3824 bp in length
* 177409 177508: gap of unknown length
* 177509 183310: contig of 5802 bp in length
* 183311 183410: gap of unknown length
* 183411 188977: contig of 5567 bp in length
* 188978 189077: gap of unknown length
* 189078 196629: contig of 7552 bp in length
* 196630 196729: gap of unknown length
* 196730 202134: contig of 5405 bp in length
* 202135 202234: gap of unknown length
* 202235 206964: contig of 4729 bp in length
* 206965 207063: gap of unknown length
* 207064 212314: contig of 5251 bp in length
* 212315 212414: gap of unknown length
* 212415 216415: contig of 4001 bp in length
* 216416 216515: gap of unknown length
* 216516 219628: contig of 3110 bp in length
* 219629 219725: gap of unknown length
* 219726 224004: contig of 4279 bp in length
* 224005 224104: gap of unknown length
* 224105 227363: contig of 3259 bp in length
* 227364 227463: gap of unknown length
* 227464 232010: contig of 4547 bp in length
* 232011 232110: gap of unknown length
* 232111 234955: contig of 2845 bp in length
* 234956 235055: gap of unknown length
* 235056 238628: contig of 3573 bp in length
* 238629 238728: gap of unknown length
* 238729 242373: contig of 3645 bp in length
* 242374 244474: contig of 2001 bp in length
* 244475 244574: gap of unknown length
* 244575 247650: contig of 3076 bp in length
* 247651 247750: gap of unknown length
* 247751 250708: contig of 2956 bp in length
* 250709 250808: gap of unknown length
* 250809 253646: contig of 2840 bp in length
* 253647 253746: gap of unknown length
* 253747 255897: contig of 2151 bp in length
* 255898 255997: gap of unknown length
* 255998 258504: contig of 2507 bp in length
* 258505 258604: gap of unknown length
* 258605 260676: contig of 2072 bp in length
* 260677 260776: gap of unknown length
* 260777 263190: contig of 2414 bp in length
* 263191 263290: gap of unknown length
* 263291 264984: contig of 1694 bp in length
* 264985 265084: gap of unknown length
* 265085 266722: contig of 1638 bp in length
* 266723 266822: gap of unknown length
* 266823 268472: contig of 1650 bp in length
* 268473 268572: gap of unknown length
* 268573 270376: contig of 1804 bp in length
* 270377 270476: gap of unknown length
* 270477 272058: contig of 1582 bp in length
* 272059 272158: gap of unknown length
* 272159 274242: contig of 2084 bp in length
* 274243 274342: gap of unknown length
* 274343 276862: contig of 2520 bp in length
* 276863 276962: gap of unknown length
* 276963 279114: contig of 2152 bp in length
* 279115 279214: gap of unknown length
* 279215 280875: contig of 1661 bp in length
* 280876 280975: gap of unknown length
* 280976 282745: contig of 1769 bp in length
* 282746 282844: gap of unknown length
* 282845 283991: contig of 1147 bp in length
* 283992 284091: gap of unknown length
* 284092 285947: contig of 1856 bp in length
* 285948 286047: gap of unknown length
* 286048 287985: contig of 1938 bp in length
* 287986 288085: gap of unknown length

288086 289898: contig of 1813 bp in length
289899 289999: gap of unknown length
290000 291415: contig of 1417 bp in length
291416 291515: gap of unknown length
291516 292676: contig of 1161 bp in length
292677 292776: gap of unknown length
292777 294223: contig of 1447 bp in length
294224 294323: gap of unknown length
294324 295422: contig of 1099 bp in length
295423 295522: gap of unknown length
295523 296939: contig of 1417 bp in length
296940 297039: gap of unknown length
297040 298270: contig of 1231 bp in length
298271 298370: gap of unknown length
298371 299936: contig of 1566 bp in length
299937 300036: gap of unknown length
300037 301414: contig of 1378 bp in length
301415 301514: gap of unknown length
301515 303105: contig of 1591 bp in length
303106 303205: gap of unknown length
303206 304965: contig of 1760 bp in length
304966 305065: gap of unknown length
305066 306362: contig of 1297 bp in length
306363 306462: gap of unknown length
306463 307463: contig of 1001 bp in length
307464 307563: gap of unknown length
307564 308714: contig of 1151 bp in length
308715 308814: gap of unknown length
308815 309956: contig of 1142 bp in length
309957 310056: gap of unknown length
310057 311344: contig of 1288 bp in length
311345 311444: gap of unknown length
311445 312654: contig of 1210 bp in length
312655 312754: gap of unknown length
312755 313811: contig of 1057 bp in length
313812 313911: gap of unknown length
313912 315044: contig of 1133 bp in length
315045 315144: gap of unknown length
315145 316580: contig of 1436 bp in length
316581 316680: gap of unknown length
316681 317778: contig of 1098 bp in length
317779 318930: contig of 1052 bp in length
318931 319030: contig of 1052 bp in length

FEATURES
      source
      1. 318930
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="RP23-345K4"

BASE COUNT  84543 a 73358 c 71912 g 82767 t 6350 others

Query Match      27.9%; Score 167; DB 2; Length 318930;
Best Local Similarity 72.7%; Pred. No. 1.le-36;
Matches 271; Conservative 0; Mismatches 95; Indels 7; Gaps 4;

QY  84  AAACAAGGTGTTCTGCTTAAACAGAGTAAGATACACACCACCCATCCATCCCTCTCTCC 143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  258116  AAAAAGGTGTTCTGCTTAAAGCAGACAGATACACACCACCCATATATATCCC-TCCTTCA 258174

QY  144  CTGTTCCCTCCCAACTTGAGTTGTTCATTCGACACAGTGTCTCTGGGTGGTAGGATGC 203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  258175  TCTGTCTCCACCCTCTCCAGTCATGTCCTCTTACAGTGTCTTGGGACCAGGAGTGC 258234

QY  204  TACAGCCACTAAGCAGGAGGAGCCCTGGAGGTGGAGGGCTTGATGTTAAGCACACC 263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  258235  TTGAACCACTAGGTTAGGGG- - - -GAGACTCGGGGAGCCCTTGTGTGATTAGCATGCT 258290

QY  264  AGAAGTGAAGCGCAAAAGGTCAGTGTCTTCATCTAGAACTCTAGATGTTCTTCCAG 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  258291  AGAACTGCAGACGCTGA-GGTCAGTGTCTTCCGCGAGAACTCTCTGGGTGTTCTTCCAG 258349

QY  324  AAAGCATCCCCGATGATCGCAGTGCAGGCACTGGCTTGTCTCTCGGTGCTACT 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


* 42240 42339: gap of unknown length
* 42340 44798: contig of 2459 bp in length
* 44799 44998: gap of unknown length
* 44899 47992: contig of 3094 bp in length
* 47993 48092: gap of unknown length
* 48093 51448: contig of 3356 bp in length
* 51449 51548: gap of unknown length
* 51549 54264: contig of 2716 bp in length
* 54265 54364: gap of unknown length
* 54365 57860: contig of 3496 bp in length
* 57861 57960: gap of unknown length
* 57961 60081: contig of 2121 bp in length
* 60082 60181: gap of unknown length
* 60182 63309: contig of 3128 bp in length
* 63310 63409: gap of unknown length
* 63410 66113: contig of 2704 bp in length
* 66114 66213: gap of unknown length
* 66214 69157: contig of 2944 bp in length
* 69158 69257: gap of unknown length
* 69258 72678: contig of 3420 bp in length
* 72679 72777: gap of unknown length
* 72778 76401: contig of 3624 bp in length
* 76402 76501: gap of unknown length
* 76502 79318: contig of 2817 bp in length
* 79319 79418: gap of unknown length
* 79420 82830: contig of 3412 bp in length
* 82831 82930: gap of unknown length
* 82931 84821: contig of 1891 bp in length
* 84822 84921: gap of unknown length
* 84922 89009: contig of 4088 bp in length
* 89010 89109: gap of unknown length
* 89110 92899: contig of 3790 bp in length
* 92900 92999: gap of unknown length
* 93000 97197: contig of 4198 bp in length
* 97198 97297: gap of unknown length
* 97298 100076: contig of 2779 bp in length
* 100077 100176: gap of unknown length
* 100177 102840: contig of 2664 bp in length
* 102841 102940: gap of unknown length
* 102941 108155: contig of 5215 bp in length
* 108156 108255: gap of unknown length
* 108256 112516: contig of 4261 bp in length
* 112517 112616: gap of unknown length
* 112617 115612: contig of 2996 bp in length
* 115613 115712: gap of unknown length
* 115713 118941: contig of 3229 bp in length
* 118942 119041: gap of unknown length
* 119042 122960: contig of 3919 bp in length
* 122961 123060: gap of unknown length
* 123061 126858: contig of 3798 bp in length
* 126859 126958: gap of unknown length
* 126959 132042: contig of 5084 bp in length
* 132043 132142: gap of unknown length
* 132143 137636: contig of 5494 bp in length
* 137637 137736: gap of unknown length
* 137737 142544: contig of 4808 bp in length
* 142545 142644: gap of unknown length
* 142645 148118: contig of 5474 bp in length
* 148119 148218: gap of unknown length
* 148219 152714: contig of 4496 bp in length
* 152715 152814: gap of unknown length
* 152815 158084: contig of 5270 bp in length
* 158085 158184: gap of unknown length
* 158185 162647: contig of 4463 bp in length
* 162648 162747: gap of unknown length
* 162748 170619: contig of 7872 bp in length
* 170620 170719: gap of unknown length

Query Match 13.1%; Score 78.4; DB 2; Length 202051;
Best Local Similarity 80.5%; Pred. No. 3e-11;
Matches 91; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
444 ACCTTGTGCATATCAACAGTCCCTTCTATACATCGGAGCTCATTATTATAGTTG 503

Db 80479 ACCTTTGGCACAATAACACTGTCCCTATCTGTACATATATGTTTTCATTGTGTAGTTG 80538
Qy 504 ATGTTGCAATTCAGAAAAAATCTCATTCTTGTCTGCTGNAAGAGTTCCTCTG 556
Db 80539 ATGTGGAATCGGAGAACAGAAATCATTCTTGTATGCTATGCTGTCAGAGTTCCTG 80591

Search completed: July 12, 2003, 14:05:19
Job time : 1250 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 12:59:52 ; Search time 160 Seconds

(without alignments)
8416.844 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 598

Sequence: 1 ttggaatagttctgttcta.....gggttagtcagattgttg 598

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
 - 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	99.8	598	21	AA50408
2	597	99.8	1231	22	AA157967
3	594.4	99.4	1512	23	ABV22854
4	594.4	99.4	1512	23	ABV28683
5	591.2	98.9	2159	22	AAH13744
6	577.6	96.6	2263	22	AAH15315
7	541.6	90.6	557	22	AAH33902
8	504	84.3	1780	22	AAH16378
9	447	74.7	15865	22	AAK90825

c 10	364.4	60.9	618	22	AAH07151
c 11	344.4	57.6	1103	22	AA159753
c 12	322.8	54.0	361	24	ABQ58609
c 13	259.8	43.4	390	23	ABV13617
c 14	237.4	39.7	856	22	AAH03726
c 15	202.6	33.9	287	23	ABV34731
c 16	202.6	33.9	287	23	ABV43586
c 17	164.2	27.5	290	23	ABV04448
c 18	125.6	21.0	563	22	AAH09037
c 19	117.2	19.6	545	24	ABQ58400
c 20	50.2	8.4	861	23	ABL07197
c 21	50.2	8.4	2830	23	ABL06544
c 22	50.2	8.4	2916	23	ABL07196
c 23	39.2	6.6	43795	21	AAZ92583
c 24	39	6.5	5139	21	AAA70139
c 25	38.6	6.5	1563	21	AAC45962
c 26	37.2	6.2	7078	23	ABL03681
c 27	37.2	6.2	13085	23	ABL03680
c 28	37	6.2	580073	18	AAV58840
c 29	36.6	6.1	104	22	ABA72131
c 30	36.6	6.1	597	22	ABA59589
c 31	36.4	6.1	21729	23	ABL06154
c 32	36.4	6.1	27423	23	ABL04268
c 33	36.2	6.1	734	22	AAH36020
c 34	35.8	6.0	4084	12	AAQ13317
c 35	35.8	6.0	4084	16	AAQ83524
c 36	35.8	6.0	4084	18	AAT72889
c 37	35.8	6.0	4084	21	AAZ98281
c 38	35.8	6.0	5282	24	AA561406
c 39	35.6	6.0	451	19	AAV57516
c 40	35.6	6.0	1963	22	ABA08826
c 41	35.6	6.0	1664976	19	AAV21209
c 42	35	5.9	3343	23	ABL07990
c 43	35	5.9	3782	23	ABL10494
c 44	35	5.9	6660	22	AAH36164
c 45	35	5.9	6660	22	AAK71438

ALIGNMENTS

RESULT 1

AA50408

ID AAA50408 standard; cdna; 598 BP.

XX AC AAA50408;

XX AC AAA50408;

XX DT 20-NOV-2000 (first entry)

XX DE Human increased in high glucose IHG-1 cdna.

XX DE IHG-1; increased in high glucose 1; human; diabetic nephropathy;

XX DE diabetes; differential expression; marker; diagnosis; ss.

XX OS Homo sapiens.

XX PN WO2000050637-A1.

XX PD 31-AUG-2000.

XX PF 28-FEB-2000; 2000WO-IE00026.

XX PR 26-FEB-1999; 99IE-0000157.

XX PA (HIBE-) HIBERGEN LTD.

XX PI (UYDU-) UNIV COLLEGE DUBLIN.

XX PI Brady HR, Godson CM, Martin FM;

XX DR WPI; 2000-572102/53.

XX PT Identifying genes used for identifying drugs for the prevention and/or therapy of diabetic nephropathy involves culturing mesangial cells in

PT the presence of glucose which induces differential expression of
 PT susceptible genes
 XX
 PS Claim 8; Page 7-8; 86pp; English.
 XX
 CC The present sequence is that of a human IHG-1 (increased in high
 CC glucose 1) cDNA partial sequence. IHG-1 was identified using a
 CC novel method for identifying genes that have a role in the
 CC presentation of diabetic nephropathy (DN). This method involves
 CC culturing mesangial cells in a medium in the presence of a
 CC concentration of glucose sufficient to induce differential
 CC expression of a gene susceptible to such differential expression.
 CC and identifying the gene so induced, especially by suppression
 CC subtractive hybridisation. Further portions of IHG-1 cDNA are
 CC given in AA50409 and AA50410. Genes identified by this method,
 CC including IHG-1, can be used as a diagnostic marker for the
 CC progression and presentation of DN, as an index of disease activity
 CC and the rate of progression of DN, and as a basis for identifying
 CC drugs for use in the prevention and/or therapy of DN (claimed).
 XX
 SQ Sequence 598 BP; 145 A; 154 C; 122 G; 176 T; 1 other;
 Query Match 99.8%; Score 597; DB 21; Length 598;
 Best Local Similarity 100.0%; Pred. No. 4.4e-174;
 Matches 598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
 DB 1 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
 QY 61 AAGGAACCATGTTCCAAACACCGCAAGGTCTTCTGCTTAAACAGAGTAAGATACACC 120
 DB 61 AAGGAACCATGTTCCAAACACCGCAAGGTCTTCTGCTTAAACAGAGTAAGATACACC 120
 QY 121 ACCCCATCATCCCTTCTTCCCTGTCCTCCCTCCCAACTTGAGTTGTGTCATTCGCACC 180
 DB 121 ACCCCATCATCCCTTCTTCCCTGTCCTCCCTCCCAACTTGAGTTGTGTCATTCGCACC 180
 QY 181 AGTGCTCTGGTGTAGGATGTCTACAGCCACCTAAGCCAGGAGCCCTGGAGGTGGGA 240
 DB 181 AGTGCTCTGGTGTAGGATGTCTACAGCCACCTAAGCCAGGAGCCCTGGAGGTGGGA 240
 QY 241 GGGCTTGCATGGTTAAGCACACCAAGTCTGCTTAAACAGAGTAAGATACACC 300
 DB 241 GGGCTTGCATGGTTAAGCACACCAAGTCTGCTTAAACAGAGTAAGATACACC 300
 QY 301 GAATCTCTGATGTTCTTCCAGAAAGCATCCCGATGATATCCAGTGCAGGGCACTG 360
 DB 301 GAATCTCTGATGTTCTTCCAGAAAGCATCCCGATGATATCCAGTGCAGGGCACTG 360
 QY 361 GCTTCTCTGTCCTGGTCCGGTCACTGCCATCTTTTCTTCCATTTCTGTGGCAGCTTAA 420
 DB 361 GCTTCTCTGTCCTGGTCCGGTCACTGCCATCTTTTCTTCCATTTCTGTGGCAGCTTAA 420
 QY 421 TTTCTTTTGTGATCACTTCAATCCACCTTCTGCGATATCAACAGTCCCTTTCTTATACA 480
 DB 421 TTTCTTTTGTGATCACTTCAATCCACCTTCTGCGATATCAACAGTCCCTTTCTTATACA 480
 QY 481 TCGGACGCTCATTTATATAGTTGATGTGAATTCAGAAACAAATCTCATTTCTGTCTG 540
 DB 481 TCGGACGCTCATTTATATAGTTGATGTGAATTCAGAAACAAATCTCATTTCTGTCTG 540
 QY 541 CTGNAAGAGTTCCTGTATCTCCCTGGCTGTACTGGTGTAGTCCAGATTGTTG 598
 DB 541 CTGNAAGAGTTCCTGTATCTCCCTGGCTGTACTGGTGTAGTCCAGATTGTTG 598
 RESULT 2
 AA157967/c
 ID AA157967 standard; cDNA; 1231 BP.
 XX
 AC AA157967;
 XX

DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 170.
 XX
 KW Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM38811.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 170; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nototropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1231 BP; 338 A; 273 C; 320 G; 300 T; 0 other;
 Query Match 99.8%; Score 597; DB 22; Length 1231;
 Best Local Similarity 99.8%; Pred. No. 6.3e-174;
 Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 60
 DB 1221 TTGGAATAGTCTTCTTTTATAAATAGTACTGCGATTAAAAAAGACACTTCTGCCA 1162
 QY 61 AAGGAACCATGTTCCAAACACCGCAAGGTGTTCTGCTTAAACAGAGTAAGATACACC 120
 DB 1161 AAGGAACCATGTTCCAAACACCGCAAGGTGTTCTGCTTAAACAGAGTAAGATACACC 1102

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 XX pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 XX 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 XX WPI; 2001-662795/76.
 DR
 XX
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer
 XX
 XX Claim 1; Page 6017; 11750pp; English.
 PS
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 CC
 XX Sequence 1512 BP; 407 A; 327 C; 401 G; 356 T; 21 other;
 SQ
 Query Match 99.4%; Score 594.4; DB 23; Length 1512;
 Best Local Similarity 99.7%; Pred. No. 4.4e-173;
 Matches 595; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 2 TGAATAGTCTTCTTTTATAAAATAGTCTCGATTAAAAAAGCACTTCTCCAA 61
 1486 TGAATAGTCTTCTTTTATAAAATAGTCTCGATTAAAAAAGCACTTCTCCAA 1427
 62 AGGAACCATGTTCCACACCGCAACAGGTGTTCTGCTTAACAGAGTAAGTACACCA 121
 1426 AGGAACCATGTTCCACACCGCAACAGGTGTTCTGCTTAACAGAGTAAGTACACCA 1367
 122 CCCCCATCCATCCCTTCCCTTCCCTCCCACTTGGTGTGTCATTGCGACCA 181
 1366 CCCCCATCCATCCCTTCCCTTCCCTCCCACTTGGTGTGTCATTGCGACCA 1307
 182 GTGTCTGGTGTAGGATGCTACAGCCACCTTAAGCGAGGAGCCCTGGAGGTGGAG 241
 1306 GTGTCTGGTGTAGGATGCTACAGCCACCTTAAGCGAGGAGCCCTGGAGGTGGAG 1247
 242 GGCTTCATGTTAAGCACACCAAGTGAAGCGCAAAAGGTCAGTCTTCATCTAG 301
 1246 GGCTTCATGTTAAGCACACCAAGTGAAGCGCAAAAGGTCAGTCTTCATCTAG 1187
 302 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGAGTGAAGGGCACTGG 361

Db 1186 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGAGTGAAGGGCACTGG 1127
 Oy 362 CTTTGTCTGGTCCGGTCACTGCATCTTTTCTTCCATTTCTGTGGCAGCTTAAT 421
 Db 1126 CTTTGTCTGGTCCGGTCACTGCATCTTTTCTTCCATTTCTGTGGCAGCTTAAT 1067
 Oy 422 TTCTTTTGTGCATCACTTTCATCCACCTTCTGCCATATCAACAGATCCCTTTCTATACAT 481
 Db 1066 TTCTTTTGTGCATCACTTTCATCCACCTTCTGCCATATCAACAGATCCCTTTCTATACAT 1007
 Oy 482 CGGCAGCTCATTTATATATAGTGTGAATTCAGAAAAAATAATCTCATTTCTGTCTGC 541
 Db 1006 CGGCAGCTCATTTATATATAGTGTGAATTCAGAAAAAATAATCTCATTTCTGTCTGC 947
 Oy 542 TGNAAAGAGTTCCTGTATATCTCCCTGGCTTGACTGGTGTAGTCAGATGTTG 598
 Db 946 TCGAAGAGTTCCTGTATATCTCCCTGGCTTGACTGGTGTAGTCAGATGTTG 890
 RESULT 5
 AAH13744/c
 ID AAH13744 standard; cDNA; 2159 BP.
 XX
 AC AAH13744;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:10648.
 XX
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 KW Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 95JP-0248036.
 PR 27-AUG-1999; 95JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PR
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10648; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 2159 BP; 607 A; 463 C; 522 G; 567 T; 0 other;

Query Match 98.9%; Score 591.2; DB 22; Length 2159;
 Best Local Similarity 99.3%; Pred. No. 5.1e-172;
 Matches 593; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TGGATAGTCTCTGTTTATATAAATAGTACTGCGATTAAATAAAAGCACTTCTGCAA 61
 DB |||||
 QY 1787 TGGATAGTCTCTGTTTATATAAATAGTACTGCGATTAAATAAAAGCACTTCTGCAA 1728
 DB |||||
 QY 62 AGGAACCATGTTTCCAAACCGCAACCAAGGTTTCTGCTTAAACAGAGTAAATACACCA 121
 DB |||||
 QY 1727 AGGAACCATGTTTCCAAACCGCAACCAAGGTTTCTGCTTAAACAGATAGATACACCA 1668
 DB |||||
 QY 122 CCCCCATCCATCCCTTCTCCCTGTTCCCTCCCACTGAGTTGTGTCATTCGCACCA 181
 DB |||||
 QY 1667 CCCCCATCCATCCCTTCTCCCTGTTCCCTCCCACTGAGTTGTGTCATTCGCACCA 1608
 DB |||||
 QY 182 GTGTCTCGGTGGTAGGATCTACAGCCACCTAAGGCAAGGAGCCCTGGAGTGGGAG 241
 DB |||||
 QY 1607 GTGTCTCGGTGGTAGGATCTACAGCCACCTAAGGCAAGGAGCCCTGGAGTGGGAG 1548
 DB |||||
 QY 242 GGCTTGATGTTTAAAGCACACCAAGTAAAGCGCAAAAGGTCAGTCTTCTATCTAG 301
 DB |||||
 QY 1547 GGCTTGATGTTTAAAGCACACCAAGTAAAGCGCAAAAGGTCAGTCTTCTATCTAG 1488
 DB |||||
 QY 302 AATCTTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGCACTGG 361
 DB |||||
 QY 1487 AATCTTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGCACTGG 1428
 DB |||||
 QY 362 CTTTGTCTGTGTCGGTCACTGCATCTTTTTCCTTCCATTTCTGTCAGCTTAAT 421
 DB |||||
 QY 1427 CTTTGTCTGTGTCGGTCACTGCATCTTTTTCCTTCCATTTCTGTCAGCTTAAT 1368
 DB |||||
 QY 422 TTTCTTTTGTATCATCTTATCCACCTTCTGTCATATCAACACAGTCCCTTTCCTATACAT 481
 DB |||||
 QY 1367 TTTCTTTTGTATCATCTTATCCACCTTCTGTCATATCAACACAGTCCCTTTCCTATACAT 1308
 DB |||||
 QY 482 CGGAGCTCATTTATATAGTTGATGTTGAATTCAGAAACAAATCTATCTTGTCTGC 541
 DB |||||
 QY 1307 CGGCGGCTCATTTATATAGTTGATGTTGAATTCAGAAACAAATCTATCTTGTCTGC 1248
 DB |||||
 QY 542 TGAAGAGTCCCGTAAATCTCCCTGGGCTTGTACTGGTGTAGTCCAGATTGTTG 598
 DB |||||
 QY 1247 TGAAGAGTCCCGTAAATCTCCCTGGGCTTGTACTGGTGTAGTCCAGATTGTTG 1191
 DB |||||

RESULT 6

AAH15315/c

ID AAH15315 standard; cDNA; 2263 BP.

XX AAH15315;

AC AAH15315;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:13474.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX

PD

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 8; SEQ ID 13474; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX AAB95893 represent human amino acid sequences; and AAH13629 to

XX CC represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

XX Sequence 2263 BP; 624 A; 486 C; 580 G; 573 T; 0 other;

SQ Query Match 96.6%; Score 577.6; DB 22; Length 2263;

Best Local Similarity 98.8%; Pred. No. 8.4e-168;

Matches 592; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 2 TGAATAGTCTTGTCTTTTATAAATAATAGTACTGCGATTAAATAAAAGCACTTCTGCAA 61

DB |||||

DB 1184 TGAATAGTCTTGTCTTTTATAAATAATAGTACTGCGATTAAATAAAAGCACTTCTGCAA 1125

QY 62 AGGAACCATGTTCCAAACCGCAACCAAGGTTTCTGCTTAAACAGAGTAAATACACCA 121

DB |||||

DB 1124 AGGAACCATGTTCCAAACCGCAACCAAGGTTTCTGCTTAAACAGAGTAAATACACCA 1065

QY 122 CCCCCATCCATCCCTTCTCCCTGTTCCCTCCCAAC--TTGAGTGTGTGTCATTGCGCAC 179

DB |||||

DB 1064 CCCCCATCCATCCCTTCTCCCTGTTCCCTCCCAAC--TTGAGTGTGTGTCATTGCGCAC 1005

QY 180 CAGTGTCTGGTGGTAGGATGCTACAGCCACTTAAGCAAGGAGCCCTGGAGTGGG 239

DB |||||

DB 1004 CAGTGTCTGGTGGTAGGATGCTACAGCCACTTAAGCAAGGAGCCCTGGAGTGGG 945

QY 240 AGGCTTGCATGGTTAAGCACACCAAGTAAAGGTCAGGTCAGTCTTCTATCT 299

DB |||||

DB 944 AGGCTTGCATGGTTAAGCACACCAAGTAAAGGTCAGGTCAGTCTTCTATCT 885

PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isoigai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 8; SEQ ID 15323; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1780 BP; 513 A; 368 C; 435 G; 464 T; 0 other;

Query Match 84.3%; Score 504; DB 22; Length 1780;
 Best Local Similarity 93.0%; Pred. No. 4.1e-145;
 Matches 555; Conservative 0; Mismatches 1; Indels 41; Gaps 1;

QY 2 TGGAAATAGTTCTTCTTTTAAATAATAGTACTGCGATTAAAAAAGCACTTCTGCCAA 61
 DB 804 TGGAAATAGTTCTTCTTTTAAATAATAGTACTGCGATTAAAAAAGCACTTCTGCCAA 745
 QY 62 AGGAACCATGTTCCACACCGCAACAGAGTGTCTGCTTAAACAGAGTAAATACACCA 121
 DB 744 AGGAACCATGTTCCACACCGCAACAGAGTGTCTGCTTAAACAGAGTAAATACACCA 685
 QY 122 CCCCCTATCCATCCCTTCTTCCCTGTTCCTCCCTCCCAACTTGAGTGTGTCATTCGCACCA 181
 DB 684 CCCCCTATCCATCCCTTCTTCCCTGTTCCTCCCTCCCAACTTGAGTGTGTCATTCGCACCA 625
 QY 182 GTGTCCTGGTGGTAGGATGCTACAGCCACCTTAAGGCAAGGAGCCCTGGGAGTGGGAG 241
 DB 624 GTGT-----CCTGGGAGTGGGAG 606
 QY 242 GGCTTGATGGTTAAGCACACACAGCACTGAAGCCGCAAAAGGTCAGTGTCTTCATCTAG 301
 DB 605 GGCTTGATGGTTAAGCACACACAGCACTGAAGCCGCAAAAGGTCAGTGTCTTCATCTAG 546
 QY 302 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGCATGATATCGCAGTGCAGGCACTGG 361
 DB 545 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGCATGATATCGCAGTGCAGGCACTGG 486
 QY 362 CTTTGTCTCGTCCGGGTCACTGCCATCTTTTTCCTTCCATTTCTGTGGCAGCTTAAT 421

Db 485 CTTTGTCTCGTCCGGGTCACTGCCATCTTTTTCCTTCCATTTCTGTGGCAGCTTAAT 426
 QY 422 TTTCTTTGTGCATCATTCACTCCACCTTCTGCCATATCAACACAGTCCCTTTCTCTATACAT 481
 Db 425 TTTCTTTGTGCATCATTCACTCCACCTTCTGCCATATCAACACAGTCCCTTTCTCTATACAT 366
 QY 482 CGGCAGCTCAATTATATAGTTGATGTTGAATTCAGAAAAAATAATCTCTCTGTCTGC 541
 Db 365 CGGCAGCTCAATTATATAGTTGATGTTGAATTCAGAAAAAATAATCTCTCTGTCTGC 306
 QY 542 TGAAGAGTTCCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTCCAGATTGTTG 598
 Db 305 TGAAGAGTTCCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTCCAGATTGTTG 249
 RESULT 9
 AAK90825/c
 ID AAK90825 standard; DNA; 15865 BP.
 XX
 AC AAK90825;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human digestive system antigen genomic sequence SEQ ID NO: 4401.
 XX
 KW Human; digestive system antigen; Gene therapy; cancer; appendicitis;
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
 KW digestive system disorder; Meckel's diverticulum; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200155314-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01324.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232337.
PR 14-SEP-2000; 2000US-0232338.
PR 14-SEP-2000; 2000US-0232339.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
(HUMA-) HUMAN GENOME SCI. INC.
PA
Rosen CA, Barash SC, Ruben SM;
PI
WPI; 2001-5026330/55.
XX
Polynucleotides encoding digestive system antigens, useful for
diagnosing, treating, preventing and/or prognosing disorders of the
digestive system, particularly cancer and cancer metastases -
XX
Disclosure; SEQ ID NO 4401; 986pp; English.
XX
The present invention provides the protein and coding sequences of a
number of human digestive system antigens. These can be used in the
diagnosis, treatment and prevention of digestive system disorders,
including cancer, Meckel's diverticulum, bacterial or parasitic
infections, appendicitis, Hirschsprung's disease, chronic colitis or
ulcerative colitis. The present sequence is a genomic DNA fragment
encoding a digestive system antigen of the invention.
XX
Sequence 15865 BP; 4775 A; 3204 C; 3309 G; 4577 T; 0 other;
SQ
Query Match 74.7%; Score 447; DB 22; Length 15865;
Best Local Similarity 98.9%; Pred. No. 5.1e-127;
Matches 450; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 TGGATAGTCTTGTCTTTATAAAATAGTACTGCGATTAAAAAAGACACTTCTGCCAA 61
DB 14717 TGGATAGTCTTGTCTTTATAAAATAGTACTGCGATTAAAAAAGACACTTCTGCCAA 14658
QY 62 AGGAACCATGTTCCACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACCA 121
DB 14657 AGGAACCATGTTCCACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACCA 14598
QY 122 CCCCCATCCATCCCTTCTTCCCTTCCCTCCCACTTGGTGTGTTGTCATTCGCACCA 181
DB 14597 CCCCCATCCATCCCTTCTTCCCTTCCCTCCCACTTGGTGTGTTGTCATTCGCACCA 14538
QY 182 GTGTCTCTGGTGGTAGGATGCTACAGCCACCTTAAGCAAGAGCCCTGGGAGTGGGAG 241
DB 14537 GTGTCTCTGGTGGTAGGATGCTACAGCCACCTTAAGCAAGAGCCCTGGGAGTGGGAG 14478
QY 242 GGCTTGCATGGTTAAGCACACCAAGAACTGAAGCGGAAAAGGTCAGCTGTCTTCTCTAG 301
DB 14477 GGCTTGCATGGTTAAGCACACCAAGAACTGAAGCGGAAAAGGTCAGCTGTCTTCTCTAG 14418

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM40597.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 3742; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA151369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 1103 BP; 303 A; 242 C; 290 G; 268 T; 0 other;
Query Match 57.6%; Score 344.4; DB 22; Length 1103;
Best Local Similarity 81.1%; Pred. No. 6.9e-96;
Matches 486; Conservative 0; Mismatches 2; Indels 111; Gaps 3;
QY 2 TGGATATAGTCTTCTTATATAAATAGTACTGCGATTAAATAAAGAGCACTTCTGCCAA 61
DB 1098 TGGATATAGTCTTCTTATATAAATAGTACTGCGATTAAATAAAGAGCACTTCTGCCAA 1039
QY 62 AGGAACCATGTTCACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACCA 121
DB 1038 AGGAACCATGTTCACACCGCAACCAAGGTGTTCTGCTTAAACAGAGTAAGATACACCA 979
QY 122 CCCCACATCCATCCCTTCCCTGTTCCCT-CCCAACTGTAGTTGTGTCATTCGCACC 180
DB 978 CCCCACATCCATCCCTTCCCTGTTCCCTGTTCCCTGTTGTTGTGTCATTCGCACC 919
QY 181 AGTGCTCTGGTGTAGGGATGCTACAGCCACCTAAGGCAAGGCC-TCGGAGGTGGG 239
DB 918 AGTGCTCTGGTGTAGGGATGCTACAGCCACCTAAGGCAAGGCC-TCGGAGGTGGG 859
QY 240 AGGCTGTGCATGGTTAAGCACACAGAACTGAAGCGCAAAAGGTCAGCTGTTCTCATCT 299
DB 858 AGGCTGTGCA----- 849
QY 300 AGAATCTGGATGTTCTTCCAGAAAGCATFCCCGATGATATCGCAGTCAAGGGCACT 359
DB 848 -----T 848
QY 360 GGCTTTGCTGTCGGTGCAGTCCATCTTTTCTTCCATTTCTGTTGGCAGCTTA 419
DB 847 GGCTTTGCTGTCGGTGCAGTCCATCTTTTCTTCCATTTCTGTTGGCAGCTTA 788
QY 420 ATTTCTTTTGTGTCATCTTCATCACCCTTCTGCCATATCAACAGATCCCTTTCCCTATAC 479
DB 787 ATTTCTTTTGTGTCATCTTCATCACCCTTCTGCCATATCAACAGATCCCTTTCCCTATAC 728
QY 480 ATCGGAGCTCATTTATATAGTTGATGTTGAATTCAGAAAAAACAATACTCATTTCTGTCT 539
DB 727 ATCGGCGGCTCATTTATATAGTTGATGTTGAATTCAGAAAAAACAATACTCATTTCTGTCT 668
QY 540 GCTGNAAGAGTTCCTCTGTAATCTCCTTGGGCTTGCTAGTCTGCTAGTTCGATTTG 598
DB 667 GCTGCAAGAGTTCCTCTGTAATCTCCTTGGGCTTGCTAGTCTGCTAGTTCGATTTG 609
RESULT 12
ABQ58609
ID ABQ58609 standard; CDNA; 361 BP.
XX AC ABQ58609;
XX DT 02-AUG-2002 (first entry)
XX DE Human colon cancer related nucleotide sequence SEQ ID NO:2304.
XX KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX OS Homo sapiens.
XX PN WO200229086-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30732.
XX PR 02-OCT-2000; 2000US-237271P.
XX PA (FARB) BAYER CORP.
XX PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiagalingam A, Lewis ME;
XX WPI: 2002-426115/45.
XX New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell
PT or tissue type, and in antisense therapy -
XX Claim 1; Fig 1; 796pp; English.
XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridises to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence of
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.
XX Sequence 361 BP; 91 A; 102 C; 81 G; 82 T; 5 other;
Query Match 54.0%; Score 322.8; DB 24; Length 361;
Best Local Similarity 97.5%; Pred. No. 1.9e-89;
Matches 346; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
QY 31 ACTGCGATTAAAAAAGACATTTCTGCCAAAGAACCATGTTTCCAAACCGCAACAAG 90
DB 1 ACTGCGATTAAAAAAGACATTTCTGCCAAAGAACCATGTTTCCAAACCGCAACAAG 60
QY 91 GTGTTCTGCTTAACAGAGTACACACCCCATCCATCCCTTCTTCTTCTTCTTCTTCTTCC 150
DB 61 GTGTTCTGCTTAACAGAGTACACACCCCATCCATCCCTTCTTCTTCTTCTTCTTCTTCC 120
QY 151 CTTCCCAACTTGTAGTTGTGTCATTTCGACCAAGTGTCTCTGGTGTAGGATGCTACAGCC 210

Db 121 CCTCCCACTTGGTGTGATTCGACACAGTGTCTCGGTGTAGGGATGCTACAGCC 180
Qy 211 ACCTAAGGCAAGAGCCCTGGAGGTGGAGGGCTTGCATGTTAAGCACACCAAGAACTG 270
Db 181 ACCTAAGGCAAGAGCCCTGGAGGTGGAGGGCTTGCATGTTAAGCACACCAAGAACTG 240
Qy 271 AAGCGCAAAAGGGTCAGC-TGCTTTCATCTAGAAATCTCTGAGATTTCTTCCAGAAAGCA 329
Db 241 AAGCGCAAAAGGGTCAGCTTGTCTTTCATCTAGAAATCTCTGAGATTTCTTCCAGAAAGCA 300
Qy 330 TCCCGGATGATATCGCAGTG-CAGGGCAGCTGGCTTTGTCTGTCGCGGTCACT 383
Db 301 TCCCGGATGATATCGCAGTG-CAGGGCAGCTGGCTTTGTCTGTCGCGGTCACT 355

RESULT 13

ABV13617/c
ID ABV13617 standard; cDNA; 390 BP.

XX AC ABV13617;

XX 13-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 13608.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 16-MAR-2000; 2000US-189862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 2258; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

XX specification or its complement. (I) is useful for:

XX (a) assessing whether a patient is afflicted with prostate cancer;

XX (b) monitoring the progression of prostate cancer in a patient;

XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;

XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;

XX (e) selecting a composition for inhibiting prostate cancer in a patient;

XX (f) assessing the prostate cell carcinogenic potential of a compound;

XX (g) determining whether prostate cancer has metastasized in a patient;

XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;

XX (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 390 BP; 126 A; 74 C; 104 G; 86 T; 0 other;

Query Match 43.4%; Score 259.8; DB 23; Length 390;
Best Local Similarity 87.9%; Pred. No. 5.7e-70;
Matches 341; Conservative 0; Mismatches 38; Indels 9; Gaps 5;

Qy 199 GATGCTACAGCACTTAAGGCAAGAGCCCTGGGA--GGTGGAGGGCTTGCATGGTTAA 256
Db 390 GATGCTTCAACCCCTTAAGGCCAGAGCCCTTGAAGGTTGAAGGGCTTGCATGGTTAA 331
Qy 257 GCACA--CCGAACTGAAGCGCAAAAGGGTCAGCTG---TCTTCATCTAGAAATCTCTGGA 311
Db 330 GCCCCCCCAGAACTTAAGCGCAAAAGGGTCAACTGTTTTTCATCTTAAATTTCTGGAAT 271
Qy 312 TGTTCCTTCCAGAAAGCATCCCCGATGATATCGCAGTGCAAGGGCACTGGCTTTGTCCTG 371
Db 270 GTTCTCTTCCAAAAAAGCAATTCCTCCATGATATCGCAGTGCAAGGGCCCTGGCTTTGTCCTG 211
Qy 372 GTCGGGTCACTGCCATCTTTTTCCTTCATTTCTGTTGGGAGCTTAA--TTTCTTTTGT 430
Db 210 GTCGGGTCACTGCCATCTTTTTCCTTCATTTCTGTTGGGAGCTTAAATTTTCTTTTGT 151
Qy 431 CATCACTTCAT--CCACCTTCTGCATATCAACACAGTCCCTTTTCCTATACATCGGCAGCT 489
Db 150 TATCACTTCATCCCACTTCTGCCATATCAACACAGTCCCTTTTCCTATACATCGGCAGCT 91
Qy 490 CATTATTATAGTTGATGTTGAATTCAGAAAAACAAAATCTCATTTCTGTCTGCTGNAAGAG 549
Db 90 CATTATTATAGTTGATGTTGAATTCAGAAAAACAAAATCTCATTTCTGTCTGCTGCAAGAG 31
Qy 550 TTCCCTGTAAATCTCCCTTGGGCTTGTAC 577
Db 30 TTCCCTGTAAATCTCCCTTGGGCTTGTAC 3

RESULT 14

AAH03726/c
ID AAH03726 standard; cDNA; 856 BP.

XX AC AAH03726;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:561.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -

XX Claim 1; SEQ ID 561; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB98993 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 856 BP; 248 A; 183 C; 222 G; 196 T; 7 other;

Query Match 39.7%; Score 237.4; DB 22; Length 856;
Best Local Similarity 94.1%; Pred. No. 7.1e-63;
Matches 255; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 329 ATCCCCGATATCGCAGTGCAAGGCGCTTGTCTCGTCCGGGTCACTGCCAT 388
Db 856 ATCCCCGATATCCCGAGNCAAGGCACTGTTGTCTCGTCCGGGTCACTGCCAT 797
QY 389 CTTTTTTCCTTCCATTTCTGTGGCAGCTTAATTTCTTTTGTCACTTCATCCACCT 447
Db 796 CTTTTTTCCTTCCATTTCTGTGGCAGCTTAATTTCTTTTGTCACTTCATCCACCT 737
QY 448 TCTGCCATATCAACAGTCCCTTCTCTATACATCGGCAGCTCATTTATATAGTTGATGT 507
Db 736 TATGCCATATCAACAGTCCCTTCTCTATACATCGGCAGCTCATTTATATAGTTGATGT 677
QY 508 TGAATTCAGAAACAAATCTCATTTCTGTCTGTGNAAGAGTTCCCTGTAATCTCCCTT 567
Db 676 TGAATTCAGAAACAAATCTCATTTCTGTCTGTGNAAGAGTTCCCTGTAATCTCCCTT 617
QY 568 GGGCTTGACTGTGTGTAGTCCAGATTGTTG 598
Db 616 GGGCTTGACTGTGTAGTCCAGATTGTTG 586

RESULT 15

ABV34731/c
ID ABV34731 standard; cDNA; 287 BP.

XX AC ABV34731;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 34722.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WC00160860-A2.

XX PD 23-AUG-2001.

XX XX 20-FEB-2001; 2001WO-US05171.

XX XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
XX

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 7273; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ Sequence 287 BP; 97 A; 55 C; 81 G; 54 T; 0 other;

Query Match 33.9%; Score 202.6; DB 23; Length 287;
Best Local Similarity 97.0%; Pred. No. 2.3e-52;
Matches 227; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 346 AGTGCAAGGGGCACTGGCTTTGTCTGGTCCGGGTCACTGCCATCTTTTTCCTTCATTT 405
Db 287 AGTGCAAGGGGCTGGCTTTGTCTGGTCCGGGTCACTGCCATCTTTTTCCTTCATTT 228
QY 406 CTGTTGGCAGCTTAA-TTCTTTTGTCACTTCAT-CCACCTTCGCATATCAACAC 463
Db 227 CTGTTGGCAGCTTAA-TTCTTTTGTCACTTCAT-CCACCTTCGCATATCAACAC 168
QY 464 AGTCCCTTTCTTATACATCGGCAGCTCATTTATATAGTTGATGTTGAAATTCAGAAACAA 523
Db 167 AGTCCCTTTCTTATACATCGGCAGCTCATTTATATAGTTGATGTTGAAATTCAGAAACAA 108
QY 524 AATCTCATTTCTTGTCTGTGNAAGAGTTCCCTGTAATCTCCCTTGGGCTTGATC 577
Db 107 AATCTCATTTCTTGTCTGTGNAAGAGTTCCCTGTAATCTCCCTTGGGCTTGATC 54

Search completed: July 12, 2003, 13:44:15
Job time : 164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 13:38:58 ; Search time 44 Seconds
(without alignments)
4168.016 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 598

Sequence: 1 ttggaatagttctgtcttta.....gggttagtcacagattgttg 598

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48.4	8.1	7218	1	US-08-232-463-14
c 2	39.2	6.6	43795	3	US-08-742-185-101
c 3	35.8	6.0	4084	2	US-08-568-459A-1
c 4	35.8	6.0	4084	2	US-08-487-826B-1
c 5	35.8	6.0	4084	4	US-09-210-288-1
c 6	35.8	6.0	4084	6	5198347-5
c 7	35.6	6.0	451	2	US-08-943-087-3
c 8	35.4	5.9	289	4	US-09-007-005-17
c 9	35.4	5.9	289	4	US-09-244-796-17
c 10	33.4	5.6	2295	1	US-08-375-300-3
c 11	33.4	5.6	2295	3	US-09-177-431-3
c 12	33.4	5.6	2295	5	PCT-US95-16930-3
c 13	33.4	5.6	4080	1	US-08-375-300-1
c 14	33.4	5.6	4080	3	US-09-177-431-1
c 15	33.4	5.6	4080	5	PCT-US95-16930-1
c 16	33.4	5.6	4705	1	US-07-998-003A-96
c 17	33.4	5.6	4705	1	US-08-453-274B-96
c 18	33.4	5.6	4705	1	US-08-453-695A-96
c 19	33.4	5.6	4705	1	US-08-268-161A-96
c 20	33.4	5.6	4705	2	US-08-453-702A-96
c 21	33.4	5.6	4705	4	US-09-099-639-96
c 22	33.4	5.6	4705	5	PCT-US93-12588-96
c 23	33.4	5.6	4705	5	PCT-US95-08071-96
c 24	33.2	5.6	2214	6	5258502-1
c 25	33	5.5	277	4	US-09-007-005-3
c 26	33	5.5	277	4	US-09-244-796-3
c 27	33	5.5	152331	3	US-09-128-155-16

28	33	5.5	176373	3	US-09-128-155-17	Sequence 17, Appl
c 29	32.8	5.5	1055	3	US-09-215-131-3	Sequence 3, Appli
c 30	32.8	5.5	1055	3	US-09-222-734-3	Sequence 3, Appli
c 31	32.8	5.5	2268	2	US-08-890-853-1	Sequence 1, Appli
c 32	32.8	5.5	2268	2	US-09-099-125A-1	Sequence 1, Appli
c 33	32.8	5.5	2268	2	US-09-099-124A-1	Sequence 1, Appli
c 34	32.8	5.5	2268	2	US-09-197-008-1	Sequence 1, Appli
c 35	32.8	5.5	2268	4	US-09-032-476-1	Sequence 1, Appli
c 36	32.8	5.5	2268	4	US-08-890-854-1	Sequence 1, Appli
c 37	32.8	5.5	2268	4	US-09-023-324-1	Sequence 1, Appli
c 38	32.8	5.5	2271	4	US-08-910-820-8	Sequence 8, Appli
c 39	32.8	5.5	2931	4	US-09-168-629-14	Sequence 14, Appl
c 40	32.8	5.5	3966	3	US-09-215-131-1	Sequence 1, Appli
c 41	32.8	5.5	3966	3	US-09-222-734-1	Sequence 1, Appli
c 42	32.6	5.5	248	4	US-09-007-005-32	Sequence 32, Appl
c 43	32.6	5.5	248	4	US-09-244-796-32	Sequence 32, Appl
c 44	32.6	5.5	4145	4	US-09-302-620B-82	Sequence 82, Appl
c 45	32.4	5.4	787	3	US-08-961-083-97	Sequence 97, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZsgpt-F1s
; US-08-232-463-14
Query Match 8.1%; Score 48.4; DB 1; Length 7218;

US-08-568-459A-1

Query Match 6.0%; Score 35.8; DB 2; Length 4084;
Best Local Similarity 53.1%; Pred. No. 0.15;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 391 TTTTCCCTTCATTTCTGTGGCAGCTTAATTTCTTTGTGCATCATCCTTCTCATCCTTCT 450
|||||
DB 1616 TTTTCCATCACATTTTCTTTCAGTTTGTGCAATCTGTGGCAATTTCTGCACAGTAAT 1557
|||||

QY 451 GCCATATCAACACAGTCCCTTTCTTCTATACATCGCAGCTCATTTATTTAGTTGATGTTGA 510
|||||

DB 1556 CCCTTCCCATCTCGAATCCATCATATATCTCGGTTCTATATTTACCGCAACATTTA 1497
|||||

QY 511 ATTCAGAAAACAAAATCTCATTC 533
|||||

DB 1496 ATTTACAAATCCATAATAAAATTC 1474
|||||

RESULT 4

US-08-487-826B-1/c
; Sequence 1, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; US-08-487-826B-1

Query Match 6.0%; Score 35.8; DB 2; Length 4084;
Best Local Similarity 53.1%; Pred. No. 0.15;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 391 TTTTCCCTTCATTTCTGTGGCAGCTTAATTTCTTTGTGCATCATCCTTCTCATCCTTCT 450
|||||

DB 1616 TTTTCCATCACATTTTCTTTCAGTTTGTGCAATTTCTGTGGCAATTTCTGCACAGTAAT 1557
|||||

QY 451 GCCATATCAACACAGTCCCTTTCTTCTATACATCGCAGCTCATTTATTTAGTTGATGTTGA 510
|||||

DB 1556 CCCTTCCCATCTCGAATCCATCATATATCTCGGTTCTATATTTACCGCAACATTTA 1497
|||||

QY 511 ATTCAGAAAACAAAATCTCATTC 533
|||||

DB 1496 ATTTACAAATCCATAATAAAATTC 1474
|||||

RESULT 5

US-09-210-288-1/c
; Sequence 1, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; US-09-210-288-1

Query Match 6.0%; Score 35.8; DB 4; Length 4084;
Best Local Similarity 53.1%; Pred. No. 0.15;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 391 TTTTCCCTTCATTTCTGTGGCAGCTTAATTTCTTTGTGCATCATCCTTCTCATCCTTCT 450
|||||

DB 1616 TTTTCCATCACATTTTCTTTCAGTTTGTGCAATTTCTGTGGCAATTTCTGCACAGTAAT 1557
|||||

QY 451 GCCATATCAACACAGTCCCTTTCTTCTATACATCGCAGCTCATTTATTTAGTTGATGTTGA 510
|||||

DB 1556 CCCTTCCCATCTCGAATCCATCATATATCTCGGTTCTATATTTACCGCAACATTTA 1497
|||||

	Query Match	5.9%; Score 35.4; DB 4; Length 289;
	Best Local Similarity	7.8%; Pred. NO. 0.043;
	Matches	18; Conservative 94; Mismatches 120; Indels 0; Gaps 0
Qy	358	CTGCTTTGTCCTGGTCACTGCCATCTTTTTTCTTCATTTCTGTGGCAGCT 417
Dd	234	CYGYCAVGVCTYGYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSY 175
Qy	418	TAATTTCTTTTGTCATCATCTTCATCCACCTTCTGCCATATCAACACAGTCCCTTCCCTAT 477
Dd	174	NINYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSY 115

;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/955,472
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, J. Peter
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 07917/050001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-9806
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2295 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-09-177-431-3

Query Match 5.6%; Score 33.4; DB 3; Length 2295;
Best Local Similarity 52.5%; Pred. No. 0.68;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 350 CAAGGCACTGGCTTTGCTGCTGGTCCGGGTCACTGCCATCTTTTTCCTTCCATTTCTGT 409
DB 1855 CAAGATCACCACCATACTCCAAATCAGAGTCGCTGCTTCTTCTTCCATCATCATCAT 1796
QY 410 TGGCAGCTTAATTTCTTTTGTCTCATCTTCCACCTTCTGCGCATATCAACACAGTCCC 469
DB 1795 CCTCATCATCATCTTTCATCATCTCCTCATCGGCTTCTTCTCCATCATCATCATCAT 1736
QY 470 TTTCCTATACATCGGCAGC 488
DB 1735 CGTCATCGTCATCGTCATC 1717

RESULT 12

PCT-US95-16930-3/c
;; Sequence 3, Application PC/TUS9516930
;; GENERAL INFORMATION:
;; APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE
;; TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF
;; TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECAY
;; TITLE OF INVENTION: FUNCTION
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson
;; STREET: 225 Franklin Street Suite 3100
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/16930
;; FILING DATE: 27-DEC-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/375,300
;; FILING DATE: 20-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, J. Peter
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 04020/046W01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)542-5070
;; TELEFAX: (617)542-8906

;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2295 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA
PCT-US95-16930-3

Query Match 5.6%; Score 33.4; DB 5; Length 2295;
Best Local Similarity 52.5%; Pred. No. 0.68;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 350 CAAGGCACTGGCTTTGCTGCTGGTCCGGGTCACTGCCATCTTTTTCCTTCCATTTCTGT 409
DB 1855 CAAGATCACCACCATACTCCAAATCAGAGTCGCTGCTTCTTCTTCCATCATCATCAT 1796
QY 410 TGGCAGCTTAATTTCTTTTGTCTCATCTTCCACCTTCTGCGCATATCAACACAGTCCC 469
DB 1795 CCTCATCATCATCTTTCATCATCTCCTCATCGGCTTCTTCTCCATCATCATCATCAT 1736
QY 470 TTTCCTATACATCGGCAGC 488
DB 1735 CGTCATCGTCATCGTCATC 1717

RESULT 13

US-08-375-300-1/c
;; Sequence 1, Application US/08375300
;; Patent No. 5679566
;; GENERAL INFORMATION:
;; APPLICANT: Feng, He
;; APPLICANT: Jacobson, Allan S.
;; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;; TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson
;; STREET: 225 Franklin Street Suite 3100
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/375,300
;; FILING DATE: 20-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, J. P.
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 04020/046001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)542-5070
;; TELEFAX: (617)542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4080 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA
US-08-375-300-1

Query Match 5.6%; Score 33.4; DB 1; Length 4080;
Best Local Similarity 52.5%; Pred. No. 0.96;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 13:41:28 ; Search time 114 Seconds
(without alignments)

8282.800 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 598

Sequence: 1 ttggaatagttcttctgtta.....ggtgttagtcagattgttg 598

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	597	99.8	1231	9	US-10-098-841-171
C 2	594.4	99.4	1512	9	US-10-198-846-10017
C 3	541.6	90.6	557	9	US-10-106-698-968
C 4	38	6.4	750	9	US-10-184-644-104
C 5	38	6.4	750	9	US-10-184-634-104
C 6	36.6	6.1	104	10	US-09-864-761-31778
C 7	36.6	6.1	597	10	US-09-864-761-15256
C 8	36.2	6.1	734	9	US-10-106-698-3112
C 9	36.2	6.1	2015	9	US-10-084-817-297
C 10	35.8	6.0	4084	9	US-10-153-273-1
C 11	35.8	6.0	185695	9	US-10-020-141-11
C 12	35.8	6.0	185695	9	US-10-017-721-1
C 13	35.4	5.9	432	9	US-09-918-995-8420
C 14	35.2	5.9	594	9	US-10-123-155-10
C 15	35.2	5.9	1919	9	US-09-938-842A-2718
C 16	35	5.9	6660	10	US-09-764-877-2529
C 17	33.6	5.6	401	10	US-09-960-352-7651
C 18	33.4	5.6	2239	10	US-09-925-301-227
C 19	33.4	5.6	2775	9	US-10-125-540-194

C 20	33.4	5.6	2775	10	US-09-764-870-194
C 21	33.2	5.6	264	9	US-10-083-357-80
C 22	33	5.5	4381	9	US-10-177-293-227
C 23	33	5.5	152331	9	US-10-095-407-16
C 24	33	5.5	176373	9	US-10-095-407-17
C 25	32.8	5.5	429	10	US-09-864-761-26138
C 26	32.8	5.5	566	10	US-09-864-761-9764
C 27	32.8	5.5	866	9	US-10-198-846-7248
C 28	32.8	5.5	1558	9	US-10-106-698-455
C 29	32.8	5.5	2268	9	US-10-156-610-1
C 30	32.8	5.5	2268	9	US-09-847-946A-1
C 31	32.8	5.5	2268	9	US-10-243-408-1
C 32	32.8	5.5	2271	9	US-09-844-988-8
C 33	32.8	5.5	2271	9	US-10-338-462-8
C 34	32.8	5.5	2271	10	US-09-844-908-8
C 35	32.8	5.5	2616	10	US-09-771-161A-50
C 36	32.8	5.5	2931	10	US-09-796-872-14
C 37	32.8	5.5	19820	10	US-09-764-877-2713
C 38	32.6	5.5	4145	9	US-10-138-838-82
C 39	32.6	5.5	4145	9	US-10-139-031-82
C 40	32.6	5.5	4145	9	US-10-138-905-82
C 41	32.6	5.5	4145	9	US-10-138-916-82
C 42	32.6	5.5	4145	9	US-09-976-800-82
C 43	32.6	5.5	4145	10	US-09-511-781-3
C 44	32.6	5.5	18996	9	US-10-270-333-61
C 45	32.6	5.5	1691139	9	US-10-067-514-1

ALIGNMENTS

RESULT 1

US-10-098-841-171/c
; Sequence 171, Application US/10098841
; Publication No US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 171
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(938)
US-10-098-841-171

Sequence 194, Appl
Sequence 80, Appl
Sequence 227, Appl
Sequence 16, Appl
Sequence 17, Appl
Sequence 26138, A
Sequence 9764, Ap
Sequence 7248, Ap
Sequence 455, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 50, Appl
Sequence 14, Appl
Sequence 2713, Ap
Sequence 82, Appl
Sequence 82, Appl
Sequence 82, Appl
Sequence 82, Appl
Sequence 82, Appl
Sequence 3, Appl
Sequence 61, Appl
Sequence 1, Appl

Query Match 99.8%; Score 597; DB 9; Length 1231;
Best Local Similarity 99.8%; Pred. No. 8.3e-178; Indels 0; Gaps 0;
Matches 597; Conservative 0; Mismatches 1;

Qy 1 TTGAATAGTCTTCTGCTTTATAAAATAGTACTCGATTAAAAAAGACACTTCTGCCA 60
Db 1221 TTGAATAGTCTTCTGCTTTATAAAATAGTACTCGATTAAAAAAGACACTTCTGCCA 1162

Qy 61 AAGAAACCATGTTTCAACACCGCAAAACAAAGTGTCTGCTTTAAACAGAGTAAGATACACC 120
Db 1161 AAGAAACCATGTTTCAACACCGCAAAACAAAGTGTCTGCTTTAAACAGAGTAAGATACACC 1102

Qy 121 ACCCCCATCCATCCCTTCTTCCCTGTTCCCTCCCACTTGAAGTGTCTGCTTTCGCAACC 180
Db 1101 ACCCCCATCCATCCCTTCTTCCCTGTTCCCTCCCACTTGAAGTGTCTGCTTTCGCAACC 1042

Qy 181 AGTGTCTGGGTGAGGATGCTACAGCCACCTTAAGCAAGGAGCCCTGGGAGTGGGA 240
Db 1041 AGTGTCTGGGTGAGGATGCTACAGCCACCTTAAGCAAGGAGCCCTGGGAGTGGGA 982

Qy 241 GGGCTTGATGTTAAGCACACCAAGAACTGAAGCGCAAAAGGTCAGTGTCTTCATCTA 300
Db 981 GGGCTTGATGTTAAGCACACCAAGAACTGAAGCGCAAAAGGTCAGTGTCTTCATCTA 922

Qy 301 GAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTG 360
Db 921 GAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTG 862

Qy 361 GCTTTGCTGCTGGTGGTCACTGCCATCTTTTTCCTTCCATTTCTGTTGGCAGCTTAA 420
Db 861 GCTTTGCTGCTGGTGGTCACTGCCATCTTTTTCCTTCCATTTCTGTTGGCAGCTTAA 802

Qy 421 TTTCTTTGTCATCACTTATCCACCTTCGCAATATCAACAGTCCCTTTCCTATACA 480
Db 801 TTTCTTTGTCATCACTTATCCACCTTCGCAATATCAACAGTCCCTTTCCTATACA 742

Qy 481 TCGCAGCTCAATTATTAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTG 540
Db 741 TCGCAGCTCAATTATTAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTG 682

Qy 541 CTGNAAGATGTTCCCTGTAATCTCCCTTGGGCTTGTAGTGGTTCAGATTTGTTG 598
Db 681 CTGNAAGATGTTCCCTGTAATCTCCCTTGGGCTTGTAGTGGTTCAGATTTGTTG 624

RESULT 2
US-10-198-846-10017/c
; Sequence 10017, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198.846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10017
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501,
; LOCATION: 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511,

; LOCATION: 1512
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10017

Query Match 99.4%; Score 594.4; DB 9; Length 1512;
Best Local Similarity 99.7%; Pred. No. 6.2e-177;
Matches 595; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGAATAGTCTTCTGCTTTATAAAATAGTACTCGATTAAAAAAGCACTTCTGCCAA 61
Db 1486 TGAATAGTCTTCTGCTTTATAAAATAGTACTCGATTAAAAAAGCACTTCTGCCAA 1427

Qy 62 AGAAACCATGTTTCAACACCGCAAAACAAAGTGTCTGCTTTAAACAGAGTAAGATACACCA 121
Db 1426 AGAAACCATGTTTCAACACCGCAAAACAAAGTGTCTGCTTTAAACAGAGTAAGATACACCA 1367

Qy 122 CCCCCATCCATCCCTTCTTCCCTGTTCCCTCCCACTTGAAGTGTCTTTCGCAACC 181
Db 1366 CCCCCATCCATCCCTTCTTCCCTGTTCCCTCCCACTTGAAGTGTCTTTCGCAACC 1307

Qy 182 GTGCTCTGGTGTGAGGATGCTACAGCCACCTTAAGCAAGGAGCCCTGGGAGTGGAG 241
Db 1306 GTGCTCTGGTGTGAGGATGCTACAGCCACCTTAAGCAAGGAGCCCTGGGAGTGGAG 1247

Qy 242 GGCTTGATGTTTAAAGCACACCAAGCACTGAAGCGCAAAAGGGTCAGTGTCTTCATCTAG 301
Db 1246 GGCTTGATGTTTAAAGCACACCAAGCACTGAAGCGCAAAAGGGTCAGTGTCTTCATCTAG 1187

Qy 302 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTGG 361
Db 1186 AATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGAAGGGCACTGG 1127

Qy 362 CTTTGTCTGGTCCGGTCACTGCCATCTTTTTCCTTCCATTTCTGTTGGCAGCTTAAT 421
Db 1126 CTTTGTCTGGTCCGGTCACTGCCATCTTTTTCCTTCCATTTCTGTTGGCAGCTTAAT 1067

Qy 422 TTTCTTTGTCATCACTTATCCACCTTCTGCCATATCAACAGTCCCTTTCCTATACAT 481
Db 1066 TTTCTTTGTCATCACTTATCCACCTTCTGCCATATCAACAGTCCCTTTCCTATACAT 1007

Qy 482 CGCAGCTCAATTATTAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTGC 541
Db 1006 CGCAGCTCAATTATTAGTTGATGTTGAATTCAGAAACAAATCTCATTTCTGCTGC 947

Qy 542 TGAAGAGTCTCCCTGTAATCTCCCTTGGGCTTGTAGTGGTTCAGATTTGTTG 598
Db 946 TGAAGAGTCTCCCTGTAATCTCCCTTGGGCTTGTAGTGGTTCAGATTTGTTG 890

RESULT 3
US-10-106-698-968/c
; Sequence 968, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 968
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-106-698-968

Query Match	90.6%;	Score 541.6;	DB 9;	Length 557;
Best Local Similarity	99.8%;	Pred. No. 1.6e-160;		
Matches 541;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;

QY	2	TGGAAATAGTCTTCCTTATATAAAATAGTACTGCGAATTAATAAAAGACACTTCTGCGCAA	61
DB	542	TGGAAATAGTCTTCGCTTTATATAAAATAGTACTGCGAATTAATAAAAGACACTTCTGCGCAA	483
QY	62	AGGAACCATGTTTCCAACACCCGCAAAACAGAGTGTTCTGCTTAAACAGAGTAAAGATACACCA	121
DB	482	AGGAACCATGTTTCCAACACCCGCAAAACAGAGTGTTCTGCTTAAACAGAGTAAAGATACACCA	423
QY	122	CCCCATCCATCCCTTCCTTCCTGTTCCCTCCCTCCCAACTTGAGTTGTGCTATTGCGACCA	181
DB	422	CCCCATCCATCCCTTCCTTCCTGTTCCCTCCCTCCCAACTTGAGTTGTGCTATTGCGACCA	363
QY	182	GTGTCCTGGGTGGTAGGATGCTACAGCCACCTAAGCGCAAGGAGCCCTGGGAGGTGGGAG	241
DB	362	GTGTCCTGGGTGGTAGGATGCTACAGCCACCTAAGCGCAAGGAGCCCTGGGAGGTGGGAG	303
QY	242	GGCTTGTCATGGTTAAGCACACCCAGAACTGAAGCGCAAAAGGGTCAGCTGTCTTCATCTAG	301
DB	302	GGCTTGTCATGGTTAAGCACACCCAGAACTGAAGCGCAAAAGGGTCAGCTGTCTTCATCTAG	243
QY	302	AATCTCTGGATGTTCTTCCAGAAAGCATCCCGCATGATATCGCAGTCAAGGGCACTGG	361
DB	242	AATCTCTGGATGTTCTTCCAGAAAGCATCCCGCATGATATCGCAGTCAAGGGCACTGG	183
QY	362	CTTTGTCTGGTCCGGGTCACTGCATCTTTTTCCTTCCATTTCTGTGGCAGCTTAAT	421
DB	182	CTTTGTCTGGTCCGGGTCACTGCATCTTTTTCCTTCCATTTCTGTGGCAGCTTAAT	123
QY	422	TTCTTTTGTTCATCACTTCATCCACCTTCTGCGCATATCAACACAGTCCCTTTTCCATATCAT	481
DB	122	TTCTTTTGTTCATCACTTCATCCACCTTCTGCGCATATCAACACAGTCCCTTTTCCATATCAT	63
QY	482	CGGCAGCTCATATTATATAGTTGATGTTGAAATTCAGAAAAACAAAATCTCAATCTTGTCTGC	541
DB	62	CGGCAGCTCATATTATATAGTTGATGTTGAAATTCAGAAAAACAAAATCTCAATCTTGTCTGC	3
QY	542	TG 543	
DB	2	TG 1	


```

RESULT 4
US-10-184-644-104/c
; Sequence 104, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 104
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-104

```

Query Match	6.4%; Score 38; DB 9; Length 750;
Best Local Similarity	9.0%; Pred. No. 0.18;
Matches	39; Conservative 132; Mismatches 263; Indels 0; Gaps 0

Qy	132	TCCCTTCCTCCCTGTTCCCTCCCAACTTGAGTTGTGTCATTCCGACCACTGTCCTGGG	191
Db	683	H..C..H..TY...K..HNKY..B..NSMH..H..Y..S..MST..A..NMETS...SH..SBSRA	624
Qy	192	TGTTAGGATGCTACAGCCACCTAAGCAAGGAGCCCTGGAGGTGGAGGGCTTGCATG	251
Db	623	MK...DMKS..SR..MHRMY..BTRHYGH...B..SNT...BKCCYB..TBA..DRM..K..HR..M	564
Qy	252	GTTAAGACACACCAGAAGTGAAGCGCAAAAGGTGAGTCTTCATCTAGTAATCTCTGGA	311
Db	563	B..RA..RBSDR...RCS..MNA..WNMARYTCST..C..Y...B...HNCSC..MS..Y..KCS...	504
Qy	312	TGTTCTTCAGAAAGCATCCCCCATGATCGCAGTGCAGGCACTGGCTTTGTCTCTG	371
Db	503	S..SMAWS..R..SMC..C..H..SM..MA..NDB..SRK...AGHY..ARNC...SHTN..RTBCY..	444
Qy	372	GTCGGGTCACTGCCATCTTTTTTCTTCATCTCTGTGGCAGCTTAATTTCTTTTGT	431
Db	443	...YSN..TW..ASC..C...THWST...AYV..YWC..MM..AC..SYB...DBBTTCS...H..CC	384
Qy	432	ATCACTTCACCCCTTCGCCATATCAACACAGTCCCTTTCCTATACATCGGCAGTCA	491
Db	383	..BSHYDC...BRYH...BTCY..AC..BNR..YAB..NASD..DKMBM..AS..NCA..C..CBNR..BM	324
Qy	492	TTATTATAGTTGATGTTGAATTCAGAAAAAATCTCATTTCTGTCTGTGNAAGAGTT	551
Db	323	..SCWSSH...TSCKM...M..THRRC...DB...S...CBT..T..CYRTR..NT..RC..A...HCT	264
Qy	552	CCCTGTAATCTCCC	565
Db	263	CN.N..NCY..BCCC	250

RESULT 5

US-10-184-634-104/c

Sequence 104, Application US/10184634

Publication No. US20030068684A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3430R1C217

CURRENT APPLICATION NUMBER: US/10/184,634

CURRENT FILING DATE: 2002-06-28

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 104

LENGTH: 750

TYPE: PRT

ORGANISM: Homo Sapien

US-10-184-634-104

Query Match	6.4%; Score 38; DB 9; Length 750;
Best Local Similarity	9.0%; Pred. No. 0.18;
Matches	39; Conservative 132; Mismatches 263; Indels 0; Gaps 0

Qy	132	TCCCTTCCTTCCTGTTCCCTCCCAACTTGAGTTGTGTCATTCCGACCACTGTCCTGGG	191
----	-----	--	-----

RESULT 5

```

US-10-184-634-104/c
; Sequence 104, Application US/10184634
; Publication No. US20030068684A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zenin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C217
CURRENT APPLICATION NUMBER: US/10/184,634
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 104
LENGTH: 750
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-634-104

Query Match      64%; Score 38; DB 9; Length 750;
Best Local Similarity 9.0%; Pred. No. 0.18;
Matches    39; Conservative   132; Mismatches 263; Indels    0; Gaps    0;

          132 TCCTTCCTTCCCTGTGTTCCCTCCCAACTTGAGTGTGTGATTCGCACCACGATGTCCTGGG 191

```



```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15256
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL139805.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
US-09-864-761-15256

Query Match          6.1%; Score 36.6; DB 10; Length 597;
Best Local Similarity 65.1%; Pred. No. 0.43; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 29;

Qy 379 TCACGGCAGCATCTTTTCCATTTCTGTGGCAGCTTAATTTCTTTTGTGTCATCACTT 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 TCATTCAGCATTAATACCTCTTCTTTCTTTTACTTTACTTTCTTCACTTCATCACTT 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 439 CATCCACTTCTGCCATATCAAC 461
    ||||| ||||| ||||| |||||
Db 368 CATTCACTTTTCAATTATCCAC 390
    ||||| ||||| ||||| |||||

RESULT 8
US-10-106-698-3112
; Sequence 3112, Application US/101066598
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3112
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (252)..(252)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (295)..(295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (313)..(315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (321)..(321)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (335)..(335)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (337)..(337)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (351)..(351)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (366)..(369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (435)..(435)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (443)..(443)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (449)..(449)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (452)..(452)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (469)..(469)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (474)..(476)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (484)..(484)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (491)..(491)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (526)..(527)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (537)..(537)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (552)..(552)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (564)..(565)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (578)..(578)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (581)..(581)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (589)..(589)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (619)..(619)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (640)..(640)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (644)..(644)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (668)..(668)
; OTHER INFORMATION: n equals a,t,g, or c
```



```
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 185695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-020-141-11

Query Match          6.0%; Score 35.8; DB 9; Length 185695;
Best Local Similarity 59.2%; Pred. No. 25;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 355 GCACCTGGCTTGTCTGCTGGGTCACTGCCATCTTTTCCCTTCATCTTCGTTGGCA 414
Db 77847 GAACCTGCTTTGTCAGCATCCCATAGTTTGTATATGTTGTTCATTTTGTGTC 77788
QY 415 GCTTAATTTCTTTGTCTATCATCTTCATCCACCTTCTGCCATAT 457
Db 77787 TCAAAATATAGTTTGATTTCTTTTCCCTCCCTCTGACCTAT 77745

RESULT 12
US-10-017-721-1/c
; Sequence 1, Application US/10017721
; Publication No. US20030096248A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Daley, George
; APPLICANT: Bolik, Stacey
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-003
; CURRENT APPLICATION NUMBER: US/10/017,721
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/317,033
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/330,248
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 185695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-721-1

Query Match          6.0%; Score 35.8; DB 9; Length 185695;
Best Local Similarity 59.2%; Pred. No. 25;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 355 GCACCTGGCTTGTCTGCTGGGTCACTGCCATCTTTTCCCTTCATCTTCGTTGGCA 414
Db 77847 GAACCTGCTTTGTCAGCATCCCATAGTTTGTATATGTTGTTCATTTTGTGTC 77788
QY 415 GCTTAATTTCTTTGTCTATCATCTTCATCCACCTTCTGCCATAT 457
Db 77787 TCAAAATATAGTTTGATTTCTTTTCCCTCCCTCTGACCTAT 77745

RESULT 13
US-09-918-995-8420/c
; Sequence 8420, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
```

```
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8420
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(432)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8420

Query Match          5.9%; Score 35.4; DB 9; Length 432;
Best Local Similarity 43.1%; Pred. No. 0.85;
Matches 84; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 331 CCCGATGATATGCGAGTGCAGGCGACTGGCTTGTCTGCTGGTCCGGGTCACTGCCATCT 390
Db 414 CCCNAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 355
QY 391 TTTTTCCTTCATTTCTGTTGGCAGCTTAATTTCTTTTGTCTCATCTTCACCTTCT 450
Db 354 TTTTNTTTTNTTTTNNNTCAAGAAATGTATCATGATTTTATTTAGTCTGCACCCAGTTAT 295
QY 451 GCCATATCAACACAGTCCCTTCTCTATACATCGGAGCTCATTTATTATAGTTGATGTTGA 510
Db 294 GATTTTAAAAACAGATCTCTTCAAGATAAGTTAACTTCTTAATATAAATGATTTT 235
QY 511 ATTGAGAAAAACAAA 525
Db 234 CTTTCAATAAAACAAA 220

RESULT 14
US-10-123-155-10/c
; Sequence 10, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P33301C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-10

Query Match          5.9%; Score 35.2; DB 9; Length 594;
Best Local Similarity 7.9%; Pred. No. 1.2;
Matches 36; Conservative 163; Mismatches 258; Indels 0; Gaps 0;
```

Search completed: July 12, 2003, 14:26:38
Job time : 117 secs

```
QY 116 ACACACCCCATCCATCCCTTCCTCCCTGTTCCCTCCCACTTGAGTTGTCATTC 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
565 A.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYS.S.S.SDSYSYA.SY 506
QY 176 GCACACAGTGCCTGGGTAGGATGCTACAGCCACCTAAGGCAAGGAGCCCTGGGAGG 235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 SYS.S.S.SWSYSYSSSDDY.CYCYRYHCHSDSYSYSYY.CRCCYTT.SYRYDCHYSC 446
QY 236 TGGAGGGCTTGATGGTTAAGCACACCAAGAACTGAAGCCAAAGGTCAGCTGCTTC 295
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
445 CCSDYCYYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSYSAYST 386
QY 296 ATCTAGATCTCTGGATGCTTCCTCCAGAAGCATCCCGATGATATCGCAGTCAAGGG 355
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 SSSSSSSSSYYTSTNYC.T.CC....T.MCAABCSTTTTTTTTTT..HSCC.SA..A. 326
QY 356 CACTGGCTTTGCTGCTCGGGTCACTGCATCTTTTTCTTCCTTCATTTCTGTTGCCAG 415
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYTRS..S.MYCY.YM.Y.YY.YYSY 266
QY 416 CTTAATTTCTTTGTCATCTCACTCACTCCCTCTGCGCATATCAACACAGTCCCTTCT 475
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 YCSRKT.M...TMDM.T.T..MHY.KYB.HCHTKCRAAT.MN.HTB...N..HBHB.B 206
QY 476 ATACATCGGAGCTCATTTATATAGTTGATGTTGAATTCAGAAAAACAATCTCATCTT 535
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
205 H.H..B.H.HSNS....TTS.....M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S.Y. 146
QY 536 GTCTGCTGNAAGAGTCCCTGTAATCTCCCTGGGCT 572
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
145 .CCR..H.R..R.G....S.CT.HN.B.CYRY.RNGMY 109
```

RESULT 15

```
US-09-938-842A-2718/c
; Sequence 2718, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2718
; LENGTH: 1919
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2718
```

```
Query Match 5.9%; Score 35.2; DB 9; Length 1919;
Best Local Similarity 60.4%; Pred. No. 2.4; Mismatches 0; Indels 0; Gaps 0;
Matches 58; Conservative 0;

QY 3 GGAATAGTCTTCTGCTTTATAAAAATAGTACTGCGATTAAAAAAGACATCTCTGCCAAA 62
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1502 GAACAGTCTATCTCTCTTAAGAAATGATTGAGATTAAACAAATGCGATTTCCGGTGA 1443
QY 63 GGAACCATGTTCCCAACCCCAACAGGTGTTCTG 98
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1442 GAATATATTTTTAAAGATCCAAATAAACAACAGATCAG 1407
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 13:29:13 ; Search time 1069 Seconds
(without alignments)

9059.776 Million cell updates/sec

Title: US-09-914-191-1

Perfect score: 598
Sequence: 1 ttggaatagttcttcttcta.....gggttagtcagattgttg 598

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	555	92.8	560	12	BF114624
2	499.4	83.5	874	9	AL563883
3	485.6	81.2	881	13	BI084544
4	474	79.3	479	10	BE206037
5	446	74.6	466	14	BM987297
6	440.8	73.7	577	10	AW954708

7	435.8	72.9	439	9	AA862635
8	421	70.4	427	9	AA308526
9	379.6	63.5	432	9	AA877204
10	370.8	62.0	374	14	W93044
11	364.4	60.9	618	9	AU134061
12	349.6	58.5	929	9	AL530910
13	345.8	57.8	357	14	W93172
14	344.4	57.6	365	14	BQ776022
15	328.8	55.0	1044	14	BQ073764
16	306.6	51.3	567	14	W72675
17	294	49.2	313	9	AI696392
18	293	49.0	233	9	AI572790
19	286.4	47.9	828	12	BQ073350
20	286.4	47.9	1026	11	AK015467
21	286.4	47.9	1185	11	AK016076
22	284.2	47.5	504	13	BM540149
23	280.8	47.0	885	14	BQ687021
24	277.4	46.4	1051	13	BM477704
25	273.6	45.8	600	10	AW549399
26	270.6	45.3	1181	11	AK017532
27	267	44.6	454	14	W77781
28	241.6	40.4	953	14	BQ068560
29	237.4	39.7	240	9	AI683779
30	237.4	39.7	856	9	AU118569
31	234.4	39.2	949	14	BQ957378
32	226.2	37.8	583	12	BG384349
33	226.2	37.8	598	12	BG384381
34	220	36.8	940	11	AK010876
35	219.6	36.7	974	13	BM459063
36	219.6	36.6	592	10	AV595816
37	199.4	33.3	225	10	AW087469
38	194	32.4	593	9	AJ392506
39	188.8	31.6	499	12	BE840989
40	176.4	29.5	404	10	AW296863
41	176	29.4	442	12	BE840990
42	162.6	27.2	287	9	AV103729
43	154.6	25.9	532	10	AW644038
44	145.6	24.3	375	12	BG018350
45	143.8	24.0	696	10	AV718160

ALIGNMENTS

RESULT 1	BF114624	560 bp	mRNA	linear	EST 24-OCT-2000
LOCUS	7j65d08.xl	Soares NSF F8_9W_OT_PA_P	SI Homo sapiens cDNA clone		
DEFINITION	IMAGE:3391311 3'	similar to TR:Q9V3N8	Q9V3N8	BG:DS00929.4	PROTEIN.
ACCESSION	BF114624				
VERSION	BF114624.1	GI:10984111			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 560)				
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.			
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40UP from Gibco High quality sequence stop: 425.				
FEATURES	Location/Qualifiers				
source	1..560				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3391311"				

/clone.lib="Soares NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HP8-9W pool 1:
758280-760583, 772104-774407 Soares NB4PA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bernaldo."
BASE COUNT 139 a 147 c 111 g 163 t
ORIGIN

Query Match 92.8%; Score 555; DB 12; Length 560;
Best Local Similarity 99.8%; Pred. No. 3.7e-142;
Matches 555; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGGAATAGTCTTCTGCTTTATAAATAAGTACTGCGATTAAAAAAGACACTTTCGCCA 60
Db 5 TTGGAATAGTCTTCTGCTTTATAAATAAGTACTGCGATTAAAAAAGACACTTTCGCCA 64
QY 61 AAGGAACCATGTTTCCAAACACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 120
Db 65 AAGGAACCATGTTTCCAAACACCGCAACAGAGTGTCTGCTTAAACAGAGTAAGATACACC 124
QY 121 ACCCCCATCATCCCTTCCCTTCCCTGTTCCCTCCCAACTTGAGTTGTGTCTCGCACC 180
Db 125 ACCCCCATCATCCCTTCCCTTCCCTGTTCCCTCCCAACTTGAGTTGTGTCTCGCACC 184
QY 181 AGTGTCTCGGTGTAGGGATGCTACAGCCACCTTAAGGCAAGGAGCCCTGGGAGGTGGGA 240
Db 185 AGTGTCTCGGTGTAGGGATGCTACAGCCACCTTAAGGCAAGGAGCCCTGGGAGGTGGGA 244
QY 241 GGGCTTCATGTTTAAAGCACACCAAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTA 300
Db 245 GGGCTTCATGTTTAAAGCACACCAAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTA 304
QY 301 GAATCTCTGATGTTCTTCAGAAACATCCCGATGATATCGCAGTGAAGGGCACTG 360
Db 305 GAATCTCTGATGTTCTTCAGAAACATCCCGATGATATCGCAGTGAAGGGCACTG 364
QY 361 GCTTTGTCTGCTCGGGTCACTGCCATCTTTTCTCTCCATTTCTGTTGGCAGCTTAA 420
Db 365 GCTTTGTCTGCTCGGGTCACTGCCATCTTTTCTCTCCATTTCTGTTGGCAGCTTAA 424
QY 421 TTTCTTTGTGATCATCATCTATCCACCTTCTGCCATATCAACAGAGTCCCTTTCTCTATACA 480
Db 425 TTTCTTTGTGATCATCATCTATCCACCTTCTGCCATATCAACAGAGTCCCTTTCTCTATACA 484
QY 481 TCGGCAGCTCATTTATATAGTGTGATTTGAATTCAGAAAAAATAATCTATCTTGTCTG 540
Db 485 TCGGCAGCTCATTTATATAGTGTGATTTGAATTCAGAAAAAATAATCTATCTTGTCTG 544
QY 541 CTGNAAGAGTTCCTCTG 556
Db 545 CTGCAAGAGTTCCTCTG 560

RESULT 2
AL563883
LOCUS
DEFINITION
AL563883 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODD003XJ17 3
prime, mRNA sequence.
ACCESSION
AL563883
VERSION
AL563883.1 GI:12913714

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 874)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1..874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSODD003XJ17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 206 a 224 c 178 g 252 t 14 others
ORIGIN

Query Match 83.5%; Score 499.4; DB 9; Length 874;
Best Local Similarity 96.3%; Pred. No. 7.9e-127;
Matches 498; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
QY 82 GCNAAACAAGTGTCTGCTTAAACAGAGTAAGATACACACCCCATCCCTCCCTT 141
Db 1 GCNAAACAAGTGTCTGCTTAAACAGAGTAAGATACACACCCCATCCCTCCCTT 60
QY 142 CCCTGTTCCTCCCAACTTGAGTTGTGTCATTCGCACCAAGTGTCTGGGTGGTAGGAT 201
Db 61 CCCTGTTCCTCCCAACTTGAGTTGTGTCATTCGCACCAAGTGTCTGGGTGGTAGGAT 120
QY 202 GCTACAGCCACCTTAAGGCAAGGAGCCCTGGGAGGTGGAGGGCTTGCATGGTTAAGCACA 261
Db 121 GCTACAGCCACCTTAAGGCAAGGAGCCCTGGKTTNNNTTGGCTTGCWTGGTTAAGCACA 180
QY 262 CGAAGCTGAAGCGCAAAAGGTCAGCTGTCTTCATCTAGATCTCTGGATGTCTCTTC 321
Db 181 CCAGAACTGAAGCGCAAAAGGTCAGCTGTCTTCATCTAGAACTCTCTGGATGTCTCTTC 240
QY 322 AGAAGCATCCCGATGATATCGCAGTGAAGGGCACTGCTTTGCTCGGTCCGGGTCA 381
Db 241 AGAAGCATCCCGATGATATCGCAGTGAAGGGCACTGCTTTGCTCGGTCCGGGTCA 300
QY 382 CTGCCATCTTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCAT 441
Db 301 CTGCCATCTTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCAT 360
QY 442 CCACCTTCCTGCATATCAACAGAGTCCCTTCCTATACATCGCAGCTCATTTATAGT 501
Db 361 CCACCTTCCTGCATATCAACAGAGTCCCTTCCTATACATCGCAGCTCATTTATAGT 420
QY 502 TGATGTTGAATTCAGAAAAAATAATCTCATTTCTGCTGCTGNAAGAGTTCCTCGTAATC 561
Db 421 TGATGTTGAATTCAGAAAAAATAATCTCATTTCTGCTGCTGCAAGADTTCCCTGTAATC 480
QY 562 TCCCTTGGGCTTGTACTGTTGTAGTCCAGATGTTG 598
|||||

Db 481 TCCCTGGGCTTGTTACTGGTGTAGTCCAGATTGTTG 517

RESULT 3

Bi084544

LOCUS

602869101T1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013768 3',
mRNA sequence.

ACCESSION

Bi084544

VERSION

Bi084544.1 GI:14502874

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 881)

TITLE

NIH-MGC http://mgc.nci.nih.gov/

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLCMI820 row: a column: 01

High quality sequence start: 28

High quality sequence stop: 722.

FEATURES

source

1. .881

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5013768"

/lab_host="NIH_MGC_102"

/tissue_type="epidermoid carcinoma, cell line"

/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed

by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-cDNA

synthesis kit (Stratagene) and Superscript II RT (Life

Technologies). Note: this is a NIH_MGC Library."

212 a

236 c

190 g

243 t

source

base

count

origin

query

match

best

local

similarity

98.5%;

pred. no. 5e-123;

mismatches

0;

indels

3;

gaps

3;

qy

73

TCCAACACCGCAACAAAGGTGTTCTGCTTAAACAGAGTAAGATACACACCCCATCCAT

132

db

23

TTCACACCGCAACAAAGGTGTTCTGCTTATACAGAGTAAGATACACACCCCATCCAT

82

qy

133

CCCTTCCTCCCTGTTCCCTCCCAACTGAGTTGTTGTCATTCGCACAGTGTCTGGT

192

db

83

CCCTTCCTCCCTGTTCCCTCCCAACTGAGTTGTTGTCATTCGCACAGTGTCTGGT

142

qy

193

GGTAGGATGCTACAGCACCTAAGGCAAGGAGCCCTGGAGGTGGGAGGCTTGCATGG

252

db

143

GGTAGGATGCTACAGCACCTAAGGCAAGGAGCCCTGGAGGTGGGAGGCTTGCATGG

202

qy

253

TTAAGCACACAGAACTGAAGCGCAAAAGGTGTCAGTGTCTTATCTAGAAATCTTCGAT

312

db

203

TTAAGCACACAGAACTGAAGCGCAAAAGGTGTCAGTGTCTTATCTAGAAATCTTCGAT

262

qy

313

GTTCCTTCAGAAAGCATCCCGATGATATCGAGTGCAGGGGACATGGCTTGTCTCTGG

372

db

263

GTTCCTTCAGAAAGCATCCCGATGATATCGAGTGCAGGGGACATGGCTTGTCTCTGG

322

373 TCCGGGTCACTGCCATCTT-TTTTCTTCCATTCTGTGTGGCAGCTTAATTTCTTTGTC 431
|||||
323 TCCGGGTCACTGCCATCTTATTTTCTTCCATTCTGTGTGGCAGCTTAATTTCTTTGTC 382
|||||
432 ATCACTTCATCCACCTTCTGCCATATCAACACAGTCCCTTCTCTATACATCGCGAGTCA 491
|||||
383 ATCACTTCATCCACCTTCTGCCATATCAACACAGTCCCTTCTCTATACATCGCGAGTCA 442
|||||
492 TTATT-ATAGTTGATGTT-GAATTCAGAAAACAAATCTCATTTCTTGTCTGCTGNAAGAG 549
|||||
443 TTATTATAGTTGATGTTTGAATTCAGAAAACAAATCTCATTTCTTGTCTGCTGCAAGAG 502
|||||
550 TTCCTCTGAATCCCTTGGCTTGACTGGTGTAGTCCAGATTGTTG 598
|||||
503 TTCCTCTGAATCCCTTGGCTTGACTGGTGTAGTCCAGATTGTTG 551
|||||

RESULT 4

BE206037

LOCUS

BE206037

DEFINITION

bb55G04.x1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3010614 3',

mRNA sequence.

ACCESSION

BE206037

VERSION

BE206037.1 GI:8749435

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 479)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Other ESTs: bb55G04.y1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-x@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

Seq primer: -40UP from Gibco

High quality sequence stop: 439...

FEATURES

Location/Qualifiers

1. .479

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3010614"

/tissue_type="rhabdomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;

Site_2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

115 a

131 c

97 g

136 t

origin

query

match

best

local

similarity

100.0%;

pred. no. 7.6e-120;

mismatches

0;

indels

0;

gaps

0;

qy

1

TTGGAATAGTCTTCTTTATATAAAATAGTACTGCGATTAAATAAAAGACATCTTGCCA

60

db

6

TTGGAATAGTCTTCTTTATATAAAATAGTACTGCGATTAAATAAAAGACATCTTGCCA

65


```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Ghelly,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.W.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other ESTs: THC162969
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers
1..427
/organism="Homo sapiens"
/db_xref="ATCC (inhost):112370"
/db_xref="taxon:9606"
/clone_lib="HCC cell line (metastasis to liver in mouse)
II"
/tissue_type="colon"
/cell_type="KM129M"
/cell_line="KM12C(HCC)metastasis into mouse (liver)"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 127 a 97 c 104 g 93 t 6 others
ORIGIN
Query Match 70.4%; Score 421; DB 9; Length 427;
Best Local Similarity 98.8%; Pred. No. 3e-105;
Matches 422; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 172 ATTCGACACAGTGTCTGGGTGGTAGGATGCTACAGCCACCTAAGCAGGAGCCCTGG 231
DB 427 ATTCGACACAGTGNCTGGGTGGTAGGATGCTACAGCCACCTAAGGAGGAGCCCTGG 368
QY 232 GAGGTGGAGGGCTTGATGCTTAAAGCACACACAGCACTGAAGCGCAAAAGGTCAGCTGT 291
DB 367 GAGGTGGAGGGCTTGATGCTTAAAGCACACACAGCACTGAAGCGCAAAAGGTCAGCTGT 308
QY 292 CTTTCATCTAGATCTCTGGATGTTCTTCAGAAAGCATCCCGATGATATCCAGTGCA 351
DB 307 CTTTCATCTAGATCTCTGGATGTTCTTCAGAAAGCATCCCGATGATATCCAGTGCA 248
QY 352 AGGCACTGCTTGTCTGCTGGGTCACTGCATCTTTTTCCTTCATCTTCTGTG 411
DB 247 AGGCACTGCTTGTCTGCTGGGTCACTGCATCTTTTTCCTTCATCTTCTGTG 188
QY 412 GCAGCTTAAATTTCTTTTGTCACTCACTTCATCCACCTTCTGCCATATCAACACAGTCCCTT 471

```

```

Db 187 GNAGCTTAATTTCTTTTGTGTCATCACCCTTCGCCATATCAACACAGTCCCTT 128
QY 472 TCCTATACATCGGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAAAATAATCTCAT 531
Db 127 TCCTATACATCGGCAGCTCATTTATTATAGTTGATGTTGAATTCAGAAAAAATAATCTCAT 68
QY 532 TCCTGTCTGCTGNAAGAGTTCCTCTGAATCTCCCTTGGGCTTCTACTGGTGTAGTCCAG 591
Db 67 TCTGTCTGCTGNAAGAGTTCCTCTGAATCTCCCTTGGGCTTCTACTGGTGTAGTCCAG 8
QY 592 ATTGTTG 598
Db 7 ATTGTTG-1

RESULT 9
AA877204 432 bp mRNA linear EST 25-MAR-1998
LOCUS Ob09g07.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1323228 3',
DEFINITION mRNA sequence.
ACCESSION AA877204
VERSION AA877204.1 GI:2986281
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 417.

FEATURES
source
Location/Qualifiers
1..432
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1323228"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 107 a 117 c 101 g 107 t
ORIGIN
Query Match 63.5%; Score 379.6; DB 9; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.8e-94;
Matches 407; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 2 TCGAATAGTCTTGTCTTATAAAATAGTACTCGGATTAAAAAAGACATCTTCTCCCA 61
Db 16 TCGAATAGTCTTGTCTTATAAAATAGTACTCGGATTAAAAAAGACATCTTCTCCCA 75
QY 62 AGGAACCATGTTCCAAACACCGCAACAGGTGTTCTGTCTTAAACAGAGTAAGACACCA 121

```


/clone lib="OVARC1"
/tissue type="ovary, tumor tissue"
/note="Vector: pME185FL3"
BASE COUNT 186 a 135 c 145 g 149 t 3 others
ORIGIN

Query Match 60.9%; Score 364.4; DB 9; Length 618;
Best Local Similarity 98.6%; Pred. No. 1.2e-89;
Matches 365; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 229 TGGAGGTGGAGGCTTGCATGTTAAGCACACAGAACTGAAGCGCAAAAGGTCAGC 288
Db 618 TGGNAGGTGGAGGCTTGCATGTTAAGCACACAGAACTGAAGCGCAAAAGGTCAGC 559
QY 289 TGTCTTCATCTAGAACTCTGGATGTTCCCTTCAGAAAGCATCCCGATGATATCGAGT 348
Db 558 TGTCTTCATCTAGAACTCTGGATGTTCCCTTCAGAAAGCATCCCGATGATATCGAGT 499
QY 349 GCAAGGACACTGGCTTTGCTGCTCGGGTCACTGCCATCTTTTCCCTTCCATTTCTG 408
Db 498 GCAAGGACACTGGCTTTGCTGCTCGGGTCACTGCCATCTTTTCCCTTCCATTTCTG 439
QY 409 TTGCAGCTTAATTTCTTTGTCATCACTTCCATCCACTTCTGCCATATCAACACAGTCC 468
Db 438 TTGCAGCTTAATTTCTTTGTCATCACTTCCATCCACTTCTGCCATATCAACACAGTCC 379
QY 469 CTTTCCATATACATCGGAGCTCATATATAGTTGATGTTGAATTCAGAAACAAAATCT 528
Db 378 CTTTCCATATACATCGGAGCTCATATATAGTTGATGTTGAATTCAGAAACAAAATCT 319
QY 529 CATTTCTGCTGCTGCAAGAGTTCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTC 588
Db 318 CATTTCTGCTGCTGCAAGAGTTCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTC 259
QY 589 CAGATTGTTG 598
Db 258 CAGATTGTTG 249

RESULT 12
AL530910/c
LOCUS
DEFINITION AL530910 LTI_NFL001_NBC4 929 bp mRNA linear EST 13-FEB-2001
prime, mRNA sequence.
ACCESSION AL530910
VERSION AL530910.1 GI:12794403
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 929)
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 929
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D0003YJ17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed.

by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 263 a 206 c 239 g 220 t 1 others
ORIGIN
Query Match 58.5%; Score 349.6; DB 9; Length 929;
Best Local Similarity 99.2%; Pred. No. 1.5e-85;
Matches 360; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
QY 237 GCGAGGCTTGCAT-GGTTAAGCACACCACTGAAGCGCAAAAGGTCAGTGTCTTC 295
Db 929 GCGAGGCTTGCATGGGTAAAGCACACCACTGAAGCGCAAAAGGTCAGTGTCTTC 870
QY 296 ATCTAGAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGG 355
Db 869 ATCTAGAATCTCTGGATGTTCTTCCAGAAAGCATCCCGATGATATCGCAGTGCAGGG 810
QY 356 CACTGGCTTTGCTGCTGCGGTCACTGCCATCTTTTCCCTTCCATTTCTGTGGCAG 415
Db 809 CACTGGCTTTGCTGCTGCGGTCACTGCCATCTTTTCCCTTCCATTTCTGTGGCAG 750
QY 416 CTTAATTTCTTTTGTGTCATCACTTCCACCTTCTGCCATATCAACACAGTCCCTTCT 475
Db 749 CTTAATTTCTTTTGTGTCATCACTTCCACCTTCTGCCATATCAACACAGTCCCTTCT 690
QY 476 ATACATCGGAGCTCATTTATATAGTTGATTTGAATTCAGAAACAAAATCTCATTTCT 535
Db 689 ATACATCGGAGCTCATTTATATAGTTGATTTGAATTCAGAAACAAAATCTCATTTCT 630
QY 536 GTCTGCTGNAAGAGTTCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTCCAGATTG 595
Db 629 GTCTGCTGCAAGAGTTCCTGTAATCTCCCTTGGGCTTGTACTGGTGTAGTCCAGATTG 570
QY 596 TTG 598
Db 569 TTG 567

RESULT 13
W93172/c
LOCUS
DEFINITION W93172 357 bp mRNA linear EST 25-NOV-1996
zd93f03.r1 Soares_fetal_heart_NDHH19W Homo sapiens cDNA clone
IMAGE:357053 5' similar to contains element TAR1 repetitive element
; mRNA sequence.

ACCESSION W93172
VERSION W93172.1 GI:1422325
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 357)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P., and Willson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 466 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 337.
Location/Qualifiers

Search completed: July 12, 2003, 14:23:25
Job time : 1072 secs

VERSION B0073764.1 GI:19902810
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 1044)
COMMENT NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2047 row: m column: 24
High quality sequence stop: 715.

FEATURES
source
1..1044
Location/Qualifiers
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:5806655"
/clone_lib="NIH_MGC_101"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."

BASE COUNT 250 a 228 c 288 g 236 t 2 others
ORIGIN

Query Match 55.0%; Score 328.8; DB 14; Length 1044;
Best Local Similarity 98.2%; Pred. No. 8.3e-80;
Matches 374; Conservative 0; Mismatches 3; Indels 4; Gaps 4;
Qy 220 AAGCAGCCCTGGGAGGTGGGAGGCTT-GCATGGTTAAGCACA-CCAGAACTGAAGCGCA 277
Db 890 AGGAAGCCCTGGGAGGTGGGAGGCTTGGCATGGTTAAGCACCACCCAGAACTGAAGCGCA 831
Qy 278 AAAGGGTCAGCTGTCTTCATCATAGAACTCTGGATGTTCTTCCAGAAAGCATCCCGCAT 337
Db 830 AAAGGGTCAGCTGTCTTCATCATAGAACTCTGGATG-TCTTCCAGAAAGCATCCCGCAT 772
Qy 338 GATATCGCAGTGAAGGCGCATGCTTTGTCCTGGTCCGGGTCACTGCCATCTTTTTC 397
Db 771 GATATCGCAGTGAAGGCGCATG-TTTGTCTGCTCGGGTCACTGCCATCTTTTTC 713
Qy 398 TTCCATTCTGCTGGCAGCTTAATTTCTTTTGTTCATCATCTTCACCTTCGCCATAT 457
Db 712 TTCCATTCTGCTGGCAGCTTAATTTCTTTTGTTCATCATCTTCACCTTCGCCATAT 653
Qy 458 CAACACAGTCCCTTTTCCTATACATCGGCAGCTCATATTATAGTTGATTTGAATTCAGA 517
Db 652 CAACACAGTCCCTTTTCCTATACATCGGCAGCTCATATTATAGTTGATTTGAATTCAGA 593
Qy 518 AAACAAATCTCATCTTGTCTGTGNAAGATTCCTCTGTAACTCCCTTGGGCTTGTAC 577
Db 592 AAACAAATCTCATCTTGTCTGTGCAAGAGTTCCCTGTAACTCCCTTGGGCTTGTAC 533
Qy 578 TGGTGTAGTCCAGATTGTTG 598
Db 532 TGGTGTAGTCCAGATTGTTG 512